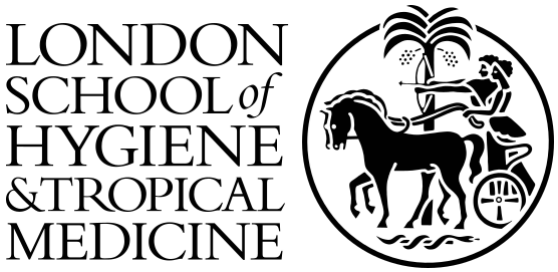


Where and for whom does the neighbourhood built environment matter for obesity and health?

Examining sources of effect heterogeneity at multiple scales in the UK adult population

KATE ELIZABETH MASON

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

December 2019

Department of Non-Communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Commonwealth Scholarship Commission in the UK

DECLARATION

I, Kate Mason, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Context: Obesity and related non-communicable diseases represent a large and growing disease burden, and various aspects of contemporary urban environments may be important drivers of these health outcomes. Neighbourhood resources for physical activity, exposure to fast-food outlets and greenspace may play a role, but the evidence regarding these neighbourhood effects remains equivocal. A possible explanation is that neighbourhood effects may be heterogeneous; some people may be more sensitive than others to the built environment, and neighbourhoods may matter more in some places than in others. We know little about how neighbourhood characteristics interact with other factors at various scales to influence health.

Methods: Using UK Biobank, a sample of >300,000 adults in mid-life, I conducted a multi-scalar examination of heterogeneity in associations between neighbourhood built environments and adiposity, CVD and cancer. I examined cross-sectional associations between characteristics of the physical activity and food environments around people's home addresses, and multiple measures of adiposity, exploring potential effect modification by individual-, neighbourhood- and local authority-level variables. Using linked hospital data, I examined effect heterogeneity in longitudinal associations with CVD and cancer.

Results: Population-wide associations obscured substantial effect heterogeneity. The magnitude of these associations varied with genetic risk, socioeconomic position, gender, and other environmental factors. Greater availability of formal physical activity facilities near home was associated with lower adiposity, particularly for people without public parks nearby, but less so for people also living near a fast-food outlet. Proximity to a fast-food outlet was associated with BMI, but this was modified by genetic risk of obesity: high-risk individuals appear to be more sensitive to their food environment. Associations between adiposity and neighbourhood characteristics varied across England, and potential drivers of this geographical heterogeneity were identified. Neighbourhood greenspace was not associated with any outcomes across the sample as a whole, but appeared protective against cancer in more deprived areas.

Conclusions: This thesis contributes empirically and conceptually to our understanding of how, where and for whom neighbourhoods matter for health. The results highlight the role of individual and contextual factors operating at multiple scales to moderate people's sensitivity to characteristics of their neighbourhood environment. A better understanding of effect modification relationships such as these may help to guide the development and evaluation of future interventions and policies involving the urban built environment.

ACKNOWLEDGEMENTS

My sincere thanks to Steve Cummins, for his expertise and time as a supervisor, and for his positivity, energy and generosity throughout, from arranging my candidature at LSHTM and helping me to develop plans and secure funding for this PhD, right through to the final stages.

Many thanks also to Neil Pearce for coming on board at short notice as supervisor, contributing valuable insights and expertise.

I also owe much gratitude to Luigi Palla, for sharing his expertise and experience, and for his generosity and thoughtfulness in providing guidance and feedback; and to Jody Phelan, for his technical expertise with genetic data.

Thanks also to Karen Lock and Matt Egan for their valuable feedback during my upgrading examination; and Frank Dudbridge and Elizabeth Williamson for advice and statistical guidance early on.

I am grateful to the Commonwealth Scholarship Commission for funding this opportunity to undertake my PhD in the UK.

I especially wish to acknowledge the work of the UK Biobank research team, including those who generated the UK Biobank Urban Morphometric Platform and other environmental data, and thank all the UK Biobank participants for their enormous contribution to this valuable resource.

Thanks also to the anonymous reviewers of the manuscripts within this thesis that have been through peer review, and to those people who provided feedback on various conference presentations I have given about this work.

Thank you to Anne Kavanagh, Rebecca Bentley and Emma Baker, among other colleagues-turned-friends in Australia for so generously supporting my career development over the years. And to new colleagues at the University of Liverpool for their encouragement and support in the final stages of submitting this thesis.

To the folk in the PHI Lab at LSHTM, for discussion, feedback, good company and moral support: thank you for making me feel a part of the group despite the fact I was rarely around. Likewise to my fellow PhD students in EPH for sharing the journey.

Thanks to my friends here in the UK for providing many welcome distractions in various forms; especially Mike for the joint study sessions and the excuse to get away from my own PhD by hanging around the edges of his, increasing my greenspace exposure and botanical vocabulary as a result.

To my loved ones in Australia who, despite being on the other side of the world, managed to give me support, encouragement and love that I felt much more closely. Thanks to you all, but in particular my mum, Maria, for always taking an interest in my work and for checking in regularly as I approached submission; my dad, John, for his thoughtful questions; and my sister, Anne, for her proofreading expertise and general good vibes. I hope this PhD partially justifies me being so far away for a little longer than expected!

And most of all, to Briony, for her constancy and support; wise insights and alternative perspective; much-appreciated pedantry, patience and suggestions on thesis drafts and half-written talks; and for sharing all the many things in life that matter more than any PhD.

ABBREVIATIONS

95% CI	95% confidence interval
BMI	Body Mass Index
CVD	Cardiovascular disease
DNA	Deoxyribonucleic acid
GLUD	Generalised Land Use Database
GP	General Practitioner
GDHI	Gross disposable household income
GWAS	Genome-wide association study
GxE	Gene-environment interaction
HES	Hospital Episode Statistics
ICD-10	International Statistical Classification of Diseases & Related Health Problems (10th edition)
LAD	Local authority district
MAF	Minor allele frequency
NCD	Non-communicable disease
NDVI	Normalised Difference Vegetation Index
NHS	National Health Service
PA	Physical activity
PC/PCA	Principal component (analysis)
RERI	Relative excess risk due to interaction
SNP	Single nucleotide polymorphism
UK	United Kingdom
UKBUMP	UK Biobank Urban Morphometric Platform
WC	waist circumference

Table of Contents

List of Tables

List of Figures

Chapter 1. BACKGROUND

1.1. Introduction

Complex health conditions such as obesity, and the major non-communicable diseases for which obesity is an important risk factor (cardiovascular disease, diabetes, cancer), represent a large and growing disease burden globally, with major implications for population health and healthcare budgets[1]. Proximate risk factors for obesity and these major NCDs include health behaviours such as excess energy intake and physical inactivity. But interventions directly targeting behaviour change are often ineffective[2] and can widen or entrench existing health inequalities by being more effective in the social groups that already experience better health[3–5]. Furthermore, the dramatic global increase in the prevalence of obesity[6] cannot plausibly be explained by a population-wide shift in personal attributes such as genetic susceptibility; instead wide-scale 'environmental' drivers of diet and physical activity behaviours must be implicated[7].

The increasing prevalence of obesity and NCDs in recent decades has occurred in parallel with global increases in urbanisation, and these are thought to be related[8,9]. The nutrition transition is closely linked to urbanisation[10], and in low- and middle-income countries, the prevalence of obesity, diabetes, CVD and cancer is typically higher in urban areas than in rural areas[11–15]. Various studies of rural-to-urban migrants in emerging economies and of migrants to Western nations have tended to show an increased risk of NCDs and their risk factors associated with such a relocation[16,17]. Various theoretically 'obesogenic' features of contemporary urban (and suburban) living – typified in Western developed countries and increasingly spreading to other settings – may at least partially explain the recent rise in obesity and related diseases[18]. Increased availability, affordability and marketing of energy-dense foods encourages less healthy diets, while increased car dependency and increasingly sedentary work and home environments encourage less active lifestyles[19,20]. With urbanisation increasing rapidly across the globe[21,22], the links between urban living and health will only become more pertinent.

Evidence of geographical inequalities in health is also abundant, at various scales including between countries[23]; regionally within countries[24]; and even within cities[25]. This lends further credence to the importance of 'place' for health, a notion long part of the public health tradition, but which fell somewhat out of favour in the latter half of the 20th century, with the rise of individualism and a focus on the lifestyle determinants of chronic disease[26]. Geographical inequalities in health are driven by a combination of contextual (i.e. place-based) and compositional (population) factors[27], and are likely to arise from a

complex interplay *between* contextual and compositional factors[28]. This complexity remains far from fully elucidated, and there is still much that is not well understood about how place and health are causally linked[29–33].

Local residential areas – usually referred to as neighbourhoods[31] – have been singled out by researchers as a geographical scale at which the built and social environments may be particularly influential for health[27], and a scale at which public health interventions might usefully be implemented. Yet despite considerable research over the past quarter of a century, the evidence base with respect to neighbourhood effects on health and health behaviours remains equivocal. A better understanding of whether, how, where and for whom these upstream, environmental contextual factors influence obesity, and NCDs such as CVD and cancer, is needed. In this thesis I seek to contribute new evidence to this area.

1.2. Socio-ecological models of health

A useful framework for thinking about these determinants is provided by socio-ecological models of health[18,34–36]. Socio-ecological models (a term I use loosely here to also include several related frameworks such as ecosocial theory[37] and eco-epidemiology[38]) recognise multiple influences on health and health behaviours, operating at multiple levels, from physiological processes within individuals up through social and environmental influences in the local sphere, to factors operating at wider macro-environmental, societal and public policy levels[39].

Socio-ecological models also often acknowledge that factors at these various levels may interact in complex ways with one another to produce health and health inequalities[40]. Such models provide a lens through which we can integrate biological, social and environmental understandings of health production[41], seek new understandings of the complexity of population health, and conceive of effective interventions to improve health and reduce health inequalities[36].

1.3. Neighbourhood built environments and health

Features of the built environment are one subset of the multiple influences on health that are represented in socio-ecological models. These may operate at levels relatively proximate to the individual, such as the residential neighbourhood, or at a wider scale such as a city, region or country. From a socio-ecological perspective, spatial and place-based features of the environments in which people live are potentially important components of the ecology of population health.

Variation in contextual factors at the neighbourhood level may be important drivers of health and health inequalities (both geographical and social). A better understanding of how such relationships work may highlight where the greatest potential for effective public health intervention lies. Early research into neighbourhood effects on health focussed on neighbourhood deprivation and poverty. Building on that, a substantial research effort has since evolved that has been focussed on specific characteristics of the built environments in which people live that are plausibly related to specific outcomes[42].

Brownson and colleagues[43] define the built environment as the "physical form of communities", including patterns of land use, built and natural features on various scales, and the transportation systems that link locations together. Various features of neighbourhood built environments have the potential to influence human health, through several mechanisms[44]. These include features that are directly toxic to human health (e.g. proximity to sources of pollution); features that constrain or facilitate healthy behaviours (e.g. access to a gym close to home making it easier to undertake regular exercise, or the relative densities of healthy and unhealthy food stores in a neighbourhood); and features that have a psychosocial impact (e.g. safe streetscapes, access to 'natural' spaces to reduce stress, and destinations that facilitate social interactions).

Considerable research attention has been paid over the past two decades to the second of these categories – built environment features that constrain or facilitate healthy behavioural choices – and in particular to the potential effects of neighbourhoods on the diet, physical activity and body weight outcomes of residents. Neighbourhood provision (or lack thereof) of opportunities to engage in adequate levels of physical activity and to sustain a healthy diet have the potential to influence energy balance (i.e. energy intake relative to energy expenditure), and therefore determine whether a person gains excess weight. With the emergence of the so-called 'obesity epidemic', obesity-related outcomes have become an obvious focal point when seeking to understand relationships between neighbourhood environments and health[42]. Reflecting the energy balance equation, neighbourhood-obesity research is typically concerned with environmental factors that may facilitate or hinder the behaviours conducive to maintaining a healthy weight: factors in either the local food environment, or the local physical activity environment, or (less often) both in combination. These have been operationalised in a multitude of ways, including both objective and perceived exposure to resources that encourage or discourage healthy diet and adequate physical activity[45].

1.3.1. An inconsistent evidence base for neighbourhood effects on health

A multitude of studies in different settings, predominantly urban contexts in developed countries, have been conducted to investigate the influence of neighbourhood characteristics on obesity and other cardiometabolic health outcomes, and on diet and physical activity behaviours. Yet equivocal and weak associations have predominated, and the evidence base to date remains inconclusive, at least in terms of clearly generalisable findings. Evidence from the USA makes it relatively clear that certain characteristics of neighbourhoods, such as land-use mix and urban sprawl, probably matter for health in that particular context. Contradictory evidence prevails in Europe and elsewhere, however, and for most neighbourhood characteristics representing the food and physical activity environments, the science remains far from settled[46–52].

The inconsistency of study findings is likely partially due to differences in methodological approaches adopted by researchers, and the degree to which the many methodological challenges of this kind of research have been adequately addressed. Several authors have noted that inconsistent findings across a wide range of environmental features may arise from a combination of factors including: underpowered studies; different study populations; inadequate adjustment for confounding; inappropriate delineation of relevant exposure areas; lack of complexity in exposure operationalisation; and the dearth of longitudinal or intervention studies to strengthen causal inference. Yet when Mackenbach and colleagues[49] identified a lack of consistency in findings and examined possible drivers of this, they found that neither methodological quality nor neighbourhood environment measurement affected the consistency of the results in the studies they reviewed. Therefore methodological differences between studies seem unlikely to fully explain heterogeneous findings across the literature.

1.4. Effect heterogeneity and the case for examining potential effect modification in neighbourhood-health research

Beyond methodological differences between studies, another possible reason for the observed inconsistency in the evidence base is true heterogeneity of effect. Effect heterogeneity is present if the strength and/or direction of an exposure-outcome relationship varies across values or strata of a third variable. That third variable is then said to be an effect modifier[53]. For example, if the effect of exercise on CVD risk is different in younger people than it is in older people, then age is acting as an effect modifier. The distribution of effect modifiers in a population affects the magnitude of the associations in that population as a whole, so the presence of effect modification has implications for

generalisability[54]. In the case of neighbourhood effects on health, if neighbourhood characteristics do not influence health and health behaviours uniformly across the population, then associations estimated in population-wide studies may mask variation in the true magnitude of the association between population subgroups. Similarly, if neighbourhood characteristics do not influence health and health behaviours uniformly across geographical space, then findings from studies in one location may not be generalisable to other settings, and studies undertaken at a national scale, for example, might mask regional differences.

In two recent papers, Keyes and Galea[55,56] argue for a "causal architecture" approach – to social epidemiology in particular – moving away from estimating global effects of a single risk factor or exposure on an outcome, and towards a greater effort to understand "the complex architectures and networks of causes that underlie disease"[56(p.2)]. They point out that "although replication of study findings increases our confidence in their validity, non-replication may be telling us something crucial about causal architecture across populations"[55]. This point has particular relevance for neighbourhood effects studies, where we observe much inconsistency of study findings. The inconsistency of study findings may point to genuinely differential effects of neighbourhood characteristics across different settings and subpopulations. Some people may be more sensitive than others to their neighbourhood environment, and neighbourhoods may matter more for health in some places than in others[27]; if so, unpacking these differences and their drivers may inform our understanding of the ways in which neighbourhood environments contribute to a complex, multi-level system of influences operating together to produce complex health outcomes such as obesity.

The possibility of effect heterogeneity also has implications for interventions, which may be less effective in some population subgroups and some locations. VanderWeele and Knol[57] explain that studying interactions between exposures (i.e. effect modification) serves an important public health function of helping to determine which of two or more subpopulations would benefit most from intervention. In their review of reviews, Ding and Gebel[50] examined the suggestions made for future research into relationships between the built environment and physical activity and/or obesity, and identified that the study of moderators (or modifiers) of these relationships was the most cited suggestion for future research. They conclude that "conceptually, it is important to identify when, where, and for whom certain environmental attributes are the most influential"[50].

Socio-ecological models of health implicitly encode interactions between factors operating at various scales, from the individual to the macro-environmental and societal [e.g. [19,39]], and in some cases they make effect modification explicit. Yet most studies that adopt a socio-ecological perspective focus on examining only the main effects of one or more of the many 'upstream' determinants of health posited in these models, rather than exploring the potentially important ways that multiple factors in a socio-ecological model interact to produce health. When interactions are examined, they are almost exclusively focussed on interactions between environments and intrapersonal characteristics, usually socio-demographic factors. Subgroup analyses (e.g. by gender or socioeconomic position) or studies focussed on more narrowly defined populations (e.g. specific ethnic groups) are not uncommon and have indeed, in some instances, revealed additional complexities that may be masked in other studies[58–61]. But these relationships are often addressed as a secondary research question, and the literature contains very limited examination of other potentially important modifiers[62].

The widely cited ANGELO framework for obesogenic environments identifies environmental influences operating at various scales (referred to as the micro and macro environments)[19]. Despite widespread recognition of multiple environmental variables as important exposures, these are rarely considered as possible modifiers of one another's effects on health. Drawing heavily on the ANGELO framework, Kremers and colleagues[63] have highlighted the conceptual importance of effect modification of the associations between environmental factors and 'energy-balance behaviours' (diet, physical activity, sedentary behaviour) but limited their consideration to individual-level modifiers of behaviour. Schneider and colleagues[64] have recently further adapted this model, but also omit a wider suite of plausible effect modifiers. Here I have extended these models further, to indicate that any of the ANGELO framework's environmental factors, along with a range of individual factors, may potentially modify the main effect of any other environmental factor on obesity or any related behaviour or health outcome (Figure 1.1).

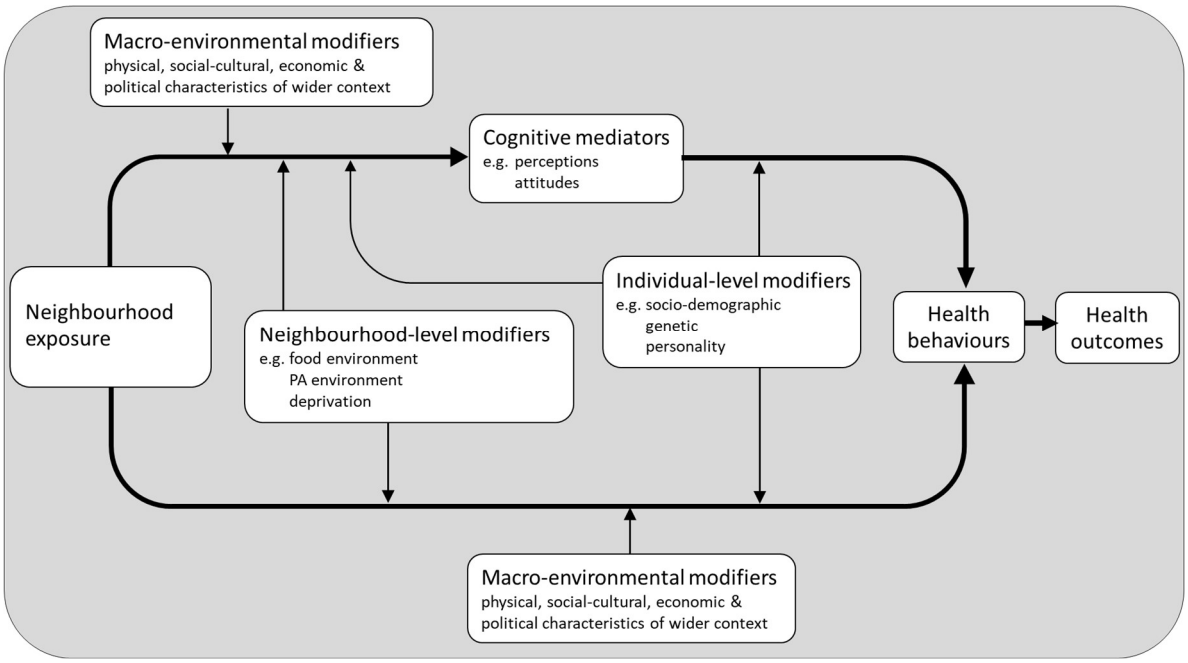


Figure 1.1 Conceptual model of relationships between neighbourhood exposures and health outcomes, modified by other factors operating at individual, neighbourhood and macro-environmental levels

Adapted from Kremers at al (2006) and Schneider et al (2017)

As Figure 1.1 indicates, the effect of any given neighbourhood exposure on health can operate directly on health behaviours such as dietary intake or physical activity (the lower pathway in Figure 1.1), or that effect can be mediated by cognitive factors such as an individual's perception of the neighbourhood, or their attitudes (e.g. towards food or exercise) (Figure 1.1, upper pathway). Health behaviours then influence health outcomes such as obesity or chronic disease risk. Importantly, both the mediated and unmediated pathways from neighbourhood exposure to health may be *modified:* by other neighbourhood-level characteristics; by individual-level factors; or by characteristics of the wider physical, socio-cultural, economic or political contexts in which people and neighbourhoods are nested (the macro environment).

Ding and Gebel's 2012 review[50] posited socio-demographic and psychosocial variables as possible individual-level effect modifiers that warrant investigation, and social environmental variables as possible neighbourhood-level modifiers. Genetic factors (e.g. genetic risk of obesity) might be added to this list: Glymour and colleagues[65] recently called for a focus on investigating genetic variation as a source of effect heterogeneity in social epidemiological studies generally, arguing that it may explain variation in sensitivity to one's environment, and similar calls have been made within the neighbourhood effects literature specifically[66]. Beyond the individual, features of the neighbourhood

environment might also interact with one another to influence health[67], and macro-environmental variables operating in the wider context in which neighbourhoods are located may also modify the influence of neighbourhood characteristics[28]. With the exception of socio-demographic variables, very few studies have examined many of these potential effect modifiers. In some cases research gaps will exist because of a paucity of appropriate data – a lack of either sufficiently large samples to enable robust analysis, or datasets that contain the relevant combination of variables. It is possible that some of the other gaps reflect publication bias if primarily null results have been generated, or they may reflect a genuine lack of research addressing those questions.

In this thesis I use a uniquely large, comprehensive and geographically diverse sample of adults in mid-life to explore heterogeneity in the associations between physical characteristics of the residential neighbourhood environment and outcome measures relating to obesity and related major NCDs. With respect to obesity-related outcomes, I focus on two exposures, each relating to one side of the energy balance equation: neighbourhood availability of formal PA facilities (energy output), and neighbourhood proximity to fast-food stores (energy input). With respect to more distal NCD outcomes, I additionally consider neighbourhood greenspace as a third exposure, which may act through pathways other than energy balance. Using the model in Figure 1.1 above as a framework, I examine potential modification of these relationships by variables operating at the individual, neighbourhood and macro-environmental levels.

Studying effect heterogeneity and its sources (effect modifiers, also referred to as moderators) in relationships between neighbourhood built environments and health can serve two important purposes: First, it can improve our understanding of the aetiology and epidemiology of important chronic health conditions, informing the development of better models of chronic disease and its many influences operating across multiple levels. Second, it can help identify the settings and subpopulations in which particular built environment interventions are likely to have the greatest public health impact, ultimately guiding more effective design of urban spaces to promote health and heath equity.

In the rest of this chapter, I will introduce the neighbourhood-health relationships around which the thesis is focused, and outline plausible primary sources of potential effect heterogeneity. I will then briefly introduce the setting of the studies contained in the thesis and the methodological approach I adopt, and outline the specific objectives of the thesis.

1.5. Neighbourhood built environments and their relationships with adiposity and NCDs

The health behaviours and outcomes on which this thesis is focused relate to obesity and the major NCDs for which obesity is an important risk factor. Specifically, I examine continuous measures of adiposity (BMI, waist circumference and percent body fat) in Chapters 4 – 7, and CVD- and cancer-related hospital admissions in Chapter 8. I investigate these outcomes in relation to various characteristics of the residential built environment.

BMI above the healthy range has been shown to be associated with most causes of death[68], and prevalence of obesity (BMI≥30 kg/m^2) has doubled in more than 70 countries across the world since 1980[69]. It is estimated that high BMI contributed to 4 million deaths and 120 million disability-adjusted life years globally in 2015 alone[69]. Excess weight defined by other measures of adiposity are also closely linked to poor health outcomes[70]. Eighteen million deaths each year are attributed to CVD globally[71] and almost 9 million to cancer, with twice that number of incident cancer cases[72].

The risk of weight gain – a result of energy imbalance – may be influenced by exposures in the residential neighbourhood environment that relate to opportunities for acquiring food and engaging in physical activity. Unhealthy diet and inadequate physical activity are major risk factors for CVD and many cancers both through and independent of weight-related pathways[73]. CVD and cancer may also be influenced by features of the built environment through pathways unrelated to diet and physical activity. Exposure to green space, for example, is hypothesised to influence risk of both CVD and some cancers through pathways involving air quality/pollution, immune function and the regulation of stress hormones[74].

If neighbourhood exposures *are* upstream determinants of these important health outcomes, a better understanding of how, where and for whom these relationships play out is important. The many equivocal and weak associations that have emerged from three decades of research on neighbourhood environments and health may indicate important effect heterogeneity that warrants further investigation.

1.6. Sources of potential effect heterogeneity

In Figure 1.1 (page 19) I categorised possible effect modifiers according to the level at which they operate: individual, neighbourhood, and macro-environment. Here I provide examples of potential modifiers at each of those levels, which I go on to investigate in this

thesis. While Figure 1.1 indicates that a modifier may act on the primary neighbourhood-health relationship at one or more of various points along the direct or mediated pathways (depicted by the multiple arrows from each modifier category), the analytical approach adopted in this thesis cannot tease apart those details. Rather, I am concerned with whether effect modification occurs at all.

1.6.1. Individual-level modifiers

While several studies of neighbourhood built environments and health have included subgroup analyses, or formally tested for effect modification by demographic factors[75–80] and a handful of other potential modifiers[81–85], it is rare in the published literature to see explicit investigation of effect modification by individual risk factors other than socio-demographic ones[62]. Certainly within the UK research has primarily been limited to testing for effect modification by gender and income[76,86,87].

1.6.1.1. Genetic risk

Most complex health conditions such as obesity and many NCDs are understood to arise from the interplay of genetic and non-genetic factors. The rapid global rise in the prevalence of these conditions in recent decades suggests an important role for changing environmental influences[88], either by direct interaction with genetic factors or by influencing behaviours that can moderate genetic risk. Obesity, for example, is known to have a heritable component[89], and the fact that not all individuals have the same behavioural or physiological responses to the near-ubiquitous changes we have seen in recent decades to our food systems, urban environments, and modern lifestyles suggests that a complex interplay between genetic and non-genetic factors affects the body mass of individuals[90,91].

The advent of genome-wide association studies (GWAS) arising from advances in genotyping technologies is enabling the genetic components of common obesity to be more readily characterised, and the wealth of new data and methods have motivated and enabled the investigation of gene-environment (GxE) interactions[90,92]. In the context of obesity, the 'environment' in GxE studies overwhelmingly refers to proximate lifestyle or behavioural factors that directly influence energy balance[93]. Only a few studies of GxE interactions have examined truly 'environmental' factors such as characteristics of the built, natural and social settings in which behavioural 'choices' are made and constrained[94–99].

The influence of neighbourhood-level exposures on BMI may vary according to level of genetic risk. The influence of the environment could be stronger among people at higher genetic risk of obesity, if genetic risk increases sensitivity to environmental factors[100,101]. Alternatively, it may be stronger among people with low genetic risk, if a healthy environment allows such people to maximise their genetic 'advantage', while people at higher genetic risk express a high-BMI phenotype regardless of external factors[93]. Any such patterns of effect heterogeneity will be obscured by examining the population as a whole without regard to underlying genetic risk, resulting in an inaccurate description of how neighbourhoods influence health[65].

1.6.1.2. Socioeconomic position

As mentioned above, where effect heterogeneity has been examined, it has often been in terms of modification by socioeconomic position. This is indeed a potentially important source of effect heterogeneity, and understanding it better may lead to more effective tailoring and targeting of built environment interventions for health. In relation to food and physical activity environments, there are several mechanisms by which socioeconomic factors might have a modifying effect. First, regardless of the availability of health-promoting resources in a neighbourhood, there may be unequal access to these resources if they require an individual to pay to access them, such as in the case of pay-to-use physical activity facilities such as gyms. Second, preferences for some neighbourhood resources may vary across socioeconomic groups, perhaps due to prevailing socio-cultural norms. For example, low-income households may be more accustomed to patronising fast-food stores because they provide inexpensive and convenient energy-dense meals[102] while social norms might encourage high-income householders to shop at high-end supermarkets[103]. Finally, effects of health-promoting neighbourhood resources that do not have prohibitive attendant costs (parks, walkable areas, low-cost leisure centres, affordable healthy food retailers) might be larger for low-income households if exposure to those resources offsets inequitable access to more costly formal physical activity facilities.

Regardless of the direction of any such effect heterogeneity, it remains a poorly understood aspect of neighbourhood-health relationships. If benefits or harms of neighbourhood characteristics depend on socioeconomic status, then any efforts to improve population health by improving neighbourhood built environments will risk widening health inequalities if they are blind to socially differential impacts.

1.6.2. Neighbourhood-level modifiers

Attempting to isolate effects of individual neighbourhood characteristics on health is likely to only ever paint an incomplete picture of how environmental factors influence the health of local residents[67,104]. Ignoring the underlying distribution of other, effect-modifying neighbourhood characteristics (beyond controlling for their possibly confounding influence), may obscure important effects in some places. While many studies have sought to characterise overall neighbourhood 'obesogenicity' through the combination of multiple neighbourhood attributes into a single composite measure[105,106] or using methods such as cluster analysis to identify neighbourhood typologies[107,108], it is surprisingly uncommon for researchers to explicitly investigate the way neighbourhood characteristics might interact to produce health effects such as obesity[62] (although some examples do exist[62,c.f.109]). It is more common for modification by neighbourhood deprivation to be studied, and numerous studies have detected evidence that this is an important modifier (e.g.[77,110]).

1.6.2.1. Built environment

With the exception of studies that examine composite measures of neighbourhood obesogenicity, most focus on the influence of specific neighbourhood exposures, largely ignoring the possibility that the effects of a given neighbourhood characteristic may not be universal, but instead vary according to other factors in the neighbourhood built environment. For example, as a potentially health-promoting neighbourhood resource[111], formal physical activity facilities may have a stronger influence on health outcomes in areas with fewer informal resources for physical activity, such as parks. Conversely, in a neighbourhood food environment dominated by fast-food stores, this competing influence on the other side of the energy balance equation might dampen the potentially health-promoting influence of physical activity facilities on body weight. Understanding whether the effect of one neighbourhood characteristic is modified by the presence of other neighbourhood characteristics, may help to identify settings in which interventions targeting a particular feature of the built environment may be more (or less) effective, and to optimise future interventions accordingly[112].

1.6.2.2. Neighbourhood deprivation

Neighbourhood deprivation may also modify effects of neighbourhood built environments. There is some (albeit mixed) evidence that residents of more deprived areas have poorer access to health-promoting neighbourhood resources – a concept know as deprivation amplification[113]. Even if resources are superficially equitably distributed, their quality (e.g. of public green spaces) may be lower in more deprived areas[113,114], and this may

cause the relationship between the neighbourhood built environment and health to vary according to area deprivation. Psychosocial stress due to neighbourhood crime or lack of social cohesion[115] may be another mechanism by which neighbourhood deprivation could act as a modifier of effects of the built environment. On the other hand, some neighbourhood resources may provide greater benefit in deprived areas if they offset other area-based inequalities (e.g.[116]).

1.6.3. Macro-environmental modifiers

Finally, evidence for a relationship between some neighbourhood characteristics and health is stronger in some geographical settings than others, and this hints at possible effect modification by factors operating at the macro-environmental scale.

Land-use mix and urban sprawl appear to have an influence on obesity-related outcomes in North America, but there is no such evidence from Europe[49]; European studies of associations between access to parks and recreation facilities and obesity yield results that are too mixed to make generalisations[49], and even within a relatively small country such as the UK, the evidence regarding those relationships is mixed[110,117,118]. Inter-city comparisons across ten countries also found variation in the association between perceived neighbourhood physical activity environments and physical activity[119]. In terms of the food environment, the evidence is dominated by studies from the United States and is equivocal overall[51,104]. From the limited number of studies conducted in the UK, greater exposure to fast-food outlets has been shown to be associated with higher BMI or greater odds of obesity in London[120], Leicester[121], Cambridgeshire[76,122] and Norfolk[123], but not in the North East of England[124] or Leeds[125,126].

Observing this kind of broad-scale geographical heterogeneity, we are compelled to ask what might be driving it, and this leads us to consider possible effect modifiers in the macro-environment. For example, might there be physical or social attributes of the wider contexts (cities, regions, nations) in which neighbourhoods are located, that cause neighbourhood environment to matter more in some settings than others? Although socio-ecological models acknowledge that these complex, multi-level relationships probably exist[19,127], and despite calls to understand 'place' as multi-scalar and relational[28], the potentially modifying roles of wider contextual factors remain underexamined.

Various macro-environmental factors have been linked to outcomes such as obesity and health behaviours such as physical activity, and any of these might also modify the influence of the neighbourhood environment on these outcomes. Such factors include

government quality and public sector expenditure[128,129], climate and weather[130,131], economic prosperity[132,133], greenspace[134], and social norms regarding health behaviours and obesity[135,136].

Geographical heterogeneity in the magnitude and direction of associations from place to place poses challenges for the generalisability of findings from many studies. Understanding the drivers of such heterogeneity may help to make sense of the inconsistent evidence base, and could ultimately have important implications for the tailoring of interventions based on local context[123].

1.7. Study setting

Understanding how contextual factors influence obesity and NCDs is a globally relevant area of public health concern, but appropriate data for large-scale observational studies of neighbourhood effects are lacking in most settings. The study of effect heterogeneity as part of this broader endeavour entails restrictive data requirements, and this is likely to be one reason it has not been extensively and routinely examined in studies of neighbourhood influences of health. There are three particular requirements that often pose a challenge.

First, larger sample sizes are required to detect interactions than are required to detect main effects. When effect modification hypotheses are not specified early in the design of a study, sample size calculations are likely to be based on achieving sufficient power to detect only main effects of a particular magnitude, leaving studies underpowered to detect interactions[137]. Good epidemiological practice demands the avoidance of underpowered and post hoc subgroup analyses[138].

Second, investigation of effect heterogeneity requires that data have been collected on effect modifiers of interest. Again, this requires *a priori* specification of effect modification hypotheses at an early stage of study design. Many neighbourhood effects studies make use of secondary data sources, and this necessarily limits researchers to the study of the potential effect modifiers for which data are available.

Third, for robust effect modification analysis, there needs to be sufficient variation in the distribution of effect modifiers across the study population and in particular across values of the primary exposure. This necessitates, again, large and diverse samples, and/or geographical diversity to ensure sufficient variation in environmental variables. In the case of genetic risk as a potential modifier, only recently has it become possible to study gene-environment interactions, as genotyping of large cohorts has become more feasible.

In the UK, where almost two in every three adults is classified as either overweight or obese[139], and where CVD accounts for 37% of deaths and cancer to a further 27%[140], there exists a comprehensive data resource, UK Biobank, with a unique breadth and depth of data that allows some of the complexity of neighbourhood-health relationships to be probed in a single very large cohort. In this PhD project I make use of data from this relatively new cohort of approximately half a million adults in mid-life. Cross-sectional data collected at recruitment to the UK Biobank cohort include detailed demographic and socioeconomic characteristics, medical history, health behaviours, psychosocial factors, and objective measurement of a range of physical, genetic and cognitive characteristics, via anthropometry, biological sampling (including genome-wide genotyping) and various other tests. UK Biobank also contains detailed and comprehensive environmental data based on objective measurement of each participant's residential neighbourhood. This dataset is described in detail in the next chapter. In addition to these extensive cross-sectional data, ongoing linkage to administrative health records for the cohort provides longitudinal data, enabling the investigation of incident health outcomes over time. Geo-location of cohort participants also allows for additional linkage by data users to external datasets, enabling the examination of relationships with place-based features of the macro-environment. The size, scope and geographical breadth of UK Biobank make it very well suited for use in the investigation of multiple potential modifiers of the associations between characteristics of neighbourhood built environments and various health outcomes.

In contrast with many studies of the neighbourhood environment and health, which tend to focus on either the general adult population or children and adolescents, this PhD is focussed on middle- to older-aged adults. This is a critical period of the lifecourse for the development of chronic disease, and is where the burden of obesity, cardiometabolic disease, many cancers, and associated healthcare costs are concentrated. This age group is also residentially more stable than some others, and potentially engages with the neighbourhood environment in ways that differ from younger or older age groups.

1.8. Thesis aims and objectives

The overarching aim of this PhD is to use this uniquely large, comprehensive and geographically diverse sample of adults in mid-life to explore heterogeneity in the associations between physical characteristics of the residential neighbourhood environment and outcome measures relating to obesity and two of the major non-communicable diseases for which obesity is a risk factor: cancer and CVD.

I address this aim by examining cross-sectional associations between characteristics of the physical activity and fast-food environments around people's home addresses and multiple measures of adiposity (BMI, waist circumference and percent body fat), exploring potential effect modification by factors operating at individual, neighbourhood and local authority levels. I then make use of linked hospital data to examine longitudinal associations between neighbourhood characteristics and CVD and cancer outcomes, including potential effect heterogeneity.

Specifically, the objectives of this PhD are:

1. To estimate the independent, cross-sectional associations between measures of the local physical activity and fast-food environments and adiposity in mid-aged adults in the UK.
2. To examine whether genetic risk of obesity modifies associations between characteristics of the neighbourhood physical activity and fast-food environments and BMI.
3. To examine whether the association between the availability of formal physical activity facilities and adiposity is modified by other physical features of the neighbourhood environment.
4. To examine possible geographical heterogeneity in the associations between neighbourhood PA and fast-food environments and BMI across England, and explore whether any such heterogeneity might be explained by locally varying contextual factors.
5. To assess whether characteristics of neighbourhood environments are associated with being admitted to hospital with cardiovascular disease or cancer, and whether these associations are modified by household income and area deprivation.

1.9. Thesis structure

This thesis is structured as a collection of five research paper manuscripts, addressing each of the primary objectives of the thesis. The first of these has already been published in *The Lancet Public Health*, the second two are being finalised for submission to *PLoS Medicine* and *Social Science and Medicine*, and the final two will be submitted to *Health and Place* and *Circulation*. Within each chapter the paper manuscript is prefaced by a cover sheet which provides information about the manuscript's publication status, copyright information, author details, and a description of my contribution to these multi-authored papers. While each manuscript is co-authored with my supervisors (and one with two additional collaborators), I am the major contributor to, and first author of, all five.

The empirical chapters (Chapters 4-8) are preceded by this introductory chapter (Chapter 1) locating the research in the various relevant literatures, describing a conceptual framework and rationale for the thesis, and outlining the thesis objectives, and by two chapters providing an overview of the data (Chapter 2) and methods (Chapter 3) used in the research. The latter complement the methods sections in the empirical chapters, which were restricted in length due to journal constraints on word counts.

To facilitate the reading of the thesis as a single, coherent body of work, I also include some linking material at the beginning of each empirical chapter. Supplementary material that did not fit within the word limits for journal submission is included in a series of appendices.

The empirical chapters are followed by a final, synthesis chapter where I provide an overview of the study findings and a broader discussion of their implications, the strengths and limitations of the analyses, directions for future research and policy, and overall conclusions.

Due to the structure of the thesis as a series of connected research papers, I present references at the end of each chapter rather than in a single list at the end of the thesis.

REFERENCES

1 Lobstein T. Prevalence and costs of obesity. *Med* 2015; **43**: 77–9.

2 Marteau T, Hollands G, Kelly M. Changing population behaviour and reducing health disparities: exploring the potential of 'choice architecture' interventions. In: Kaplan RM, Spittel ML, David DH, eds. Population Health: Behavioral and Social Science Insights. Rockville, MD: Agency for Healthcare Research and Quality and Office of Behavioral and Social Sciences Research, National Institutes of Health, 2015: 105–26.

3 Baum F, Fisher M. Why behavioural health promotion endures despite its failure to reduce health inequities. *Sociol Health Illn* 2014; **36**: 213–25.

4 Beauchamp A, Backholer K, Magliano D, Peeters A. The effect of obesity prevention interventions according to socioeconomic position: a systematic review. *Obes Rev* 2014; **15**: 541–54.

5 Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013; **67**: 190–3.

6 Ng M, Fleming T, Robinson M, *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; **384**: 766–81.

7 Rodgers A, Woodward A, Swinburn B, Dietz WH. Prevalence trends tell us what did not precipitate the US obesity epidemic. *Lancet Public Heal* 2018; **3**: e162–3.

8 Wagner KH, Brath H. A global view on the development of non communicable diseases. *Prev Med* 2012; **54**: S38-41.

9 Popkin BM. Nutrition transition and the global diabetes epidemic. *Curr Diab Rep* 2015; **15**: 64.

10 International Food Policy Research Institute. 2018 Global Food Policy Report. Washington, DC, 2018

11 Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004; **27**: 1047–53.

12 Popkin BM. The nutrition transition and its health implications in lower-income countries. *Public Health Nutr* 1998; **1**: 5–21.

13 World Health Organization. Obesity: preventing and managing the global epidemic. Geneva: WHO, 2000.

14 Stuckler D. Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations. Milbank Q 2008; **86**: 273-326.

15 Gong P, Liang S, Carlton EJ, *et al.* Urbanisation and health in China. *Lancet* 2012; **379**: 843–52.

16 Ebrahim S, Kinra S, Bowen L, *et al.* The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Med* 2010; **7**: e1000268.

17 Kaplan MS, Huguet N, Newsom JT, McFarland BH. The association between length of residence and obesity among Hispanic immigrants. *Am J Prev Med* 2004; **27**: 323–6.

18 Egger G, Swinburn B. An 'ecological' approach to the obesity pandemic. *Br Med J* 1997; **315**: 144–80.

19 Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Prev Med* 1999; **29**: 563–70.

20 Black JL, Macinko J. Neighborhoods and obesity. *Nutr Rev* 2008; **66**: 2–20.

21 United Nations. World Urbanization Prospects: The 2018 Revision. 2018. http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf

22 Seto KC, Fragkias M, Güneralp B, Reilly MK. A meta-analysis of global urban land expansion. *PLoS One* 2011; **6**: e23777.

23 World Health Organization. Global status report on noncommunicable diseases 2014. Geneva: WHO, 2014.

24 Shelton NJ. Regional risk factors for health inequalities in Scotland and England and the 'Scottish effect'. *Soc Sci Med* 2009; **69**: 761–7.

25 CSDH. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. Geneva, 2008.

26 Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002; **55**: 125–39.

27 Macintyre S, Ellaway A. Neighborhoods and health: an overview. In: Kawachi I, Berkman LF, eds. Neighborhoods and Health. Oxford: Oxford University Press, 2003: 20–42.

28 Cummins S, Curtis S, Diez-Roux A V., Macintyre S. Understanding and representing 'place' in health research: a relational approach. *Soc Sci Med* 2007; **65**: 1825–38.

29 Pearce JR, Richardson EA, Mitchell RJ, Shortt NK. Environmental justice and health: a study of multiple environmental deprivation and geographical inequalities in health in New Zealand. *Soc Sci Med* 2011; **73**: 410–20.

30 Riva M, Gauvin L, Apparicio P, Brodeur JM. Disentangling the relative influence of built and socioeconomic environments on walking: the contribution of areas homogenous along exposures of interest. *Soc Sci Med* 2009; **69**: 1296–305.

31 Diez-Roux A V, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010; **1186**: 125–45.

32 Moon G, Quarendon G, Barnard S, Twigg L, Blyth B. Fat nation: deciphering the distinctive geographies of obesity in England. *Soc Sci Med* 2007; **65**: 20–31.

33 Bambra C, Smith KE, Pearce J. Scaling up: the politics of health and place. *Soc Sci Med* 2019.

34 Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. *Soc Sci Med* 2006; **62**: 1650–71.

35 Stokols D. Establishing and maintaining healthy environments: toward a social ecology of health promotion. *Am Psychol* 1992; **47**: 6–22.

36 Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health. Background document to WHO – Strategy paper. 1991

37 Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 1994; **39**: 887–903.

38 Susser M, Susser E. Choosing a future for epidemiology II: from black box to Chinese boxes and eco-epidemiology. *Am J Public Health* 1996; **86**: 674–7.

39 Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz K, Rimer BK, Viswanath K, eds. Health Behavior and Health Education: Theory, Research, and Practice, 4th edn. San Francisco: Jossey-Bass, 2008: 465–85.

40 Dahlgren G, Whitehead M. European strategies for tackling social inequities in health: levelling up part 2. Copenhagen: World Health Organization, 2006.

41 Krieger N. Epidemiology and the people's health: theory and context. New York: Oxford University Press, 2011.

42 Duncan DT, Kawachi I. Neighborhoods and health: a progress report. In: Duncan DT, Kawachi I, eds. Neighborhoods and Health, 2nd edn. New York: Oxford University Press, 2018: 1–16.

43 Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the Built Environment for Physical Activity. State of the Science. *Am J Prev Med* 2009; **36**: S99-S123.e12.

44 Gullón P, Lovasi GS. Designing healthier built environments. In: Duncan DT, Kawachi I, eds. Neighborhoods and Health, 2nd edn. New York: Oxford University Press, 2018: 219–46.

45 Townshend T, Lake AA. Obesogenic urban form: theory, policy and practice. *Health Place* 2009; **15**: 909–16.

46 Black C, Moon G, Baird J. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health Place* 2014; **27**: 229–42.

47 Bancroft C, Joshi S, Rundle A, *et al.* Association of proximity and density of parks and objectively measured physical activity in the United States: a systematic review. *Soc Sci Med* 2015; **138**: 22–30.

48 Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW. Correlates of physical activity: why are some people physically active and others not? *Lancet* 2012; **380**: 258–71.

49 Mackenbach JD, Rutter H, Compernolle S, *et al.* Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. *BMC Public Health* 2014; **14**: 233.

50 Ding D, Gebel K. Built environment, physical activity, and obesity: what have we learned from reviewing the literature? *Heal Place* 2012; **18**: 100–5.

51 Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity* 2015; **23**: 1331–44.

52 Koohsari MJ, Mavoa S, Villanueva K, *et al.* Public open space, physical activity, urban design and public health: concepts, methods and research agenda. *Health Place* 2015; **33**: 75–82.

53 Greenland S, Rothman KJ, Lash T. Measures of effect and measures of association. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology, 3rd edn. Philadelphia, PA: Wolters Kluwer Health, 2008.

54 Hernán MA, Vanderweele TJ. Compound treatments and transportability of causal inference. *Epidemiology* 2011.

55 Keyes K, Galea S. What matters most: quantifying an epidemiology of consequence. *Ann Epidemiol* 2015; **25**: 305–11.

56 Keyes KM, Galea S. Commentary: The limits of risk factors revisited: is it time for a causal architecture approach? *Epidemiology* 2017; **28**: 1–5.

57 VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Method* 2014; **3**: 33–72.

58 Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. *Epidemiol Rev* 2009; **31**: 7–20.

59 Jones-Smith JC, Karter AJ, Warton EM, *et al.* Obesity and the food environment: income and ethnicity differences among people with diabetes: the Diabetes Study of Northern California (DISTANCE). *Diabetes Care* 2013; **36**: 2697–705.

60 Richardson EA, Mitchell R. Gender differences in relationships between urban green space and health in the United Kingdom. *Soc Sci Med* 2010; **71**: 568–75.

61 Markevych I, Schoierer J, Hartig T, *et al.* Exploring pathways linking greenspace to health: theoretical and methodological guidance. *Environ Res* 2017; **158**: 301–17.

62 Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev* 2011; **12**: 217–30.

63 Kremers SPJ, de Bruijn G-J, Visscher TLS, van Mechelen W, de Vries NK, Brug J. Environmental influences on energy balance-related behaviors: a dual-process view. *Int J Behav Nutr Phys Act* 2006; **3**: 9.

64 Schneider S, Diehl K, Görig T, *et al.* Contextual influences on physical activity and eating habits: options for action on the community level. *BMC Public Health* 2017; **17**: 760.

65 Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: Off-roading with social epidemiology: exploration, causation, translation. *Am J Epidemiol* 2013; **178**: 858–63.

66 Diez Roux A V. Conceptual approaches to the study of health disparities. *Annu Rev Public Health* 2012; **33**: 41–58.

67 Myers CA, Denstel KD, Broyles ST. The context of context: examining the associations between healthy and unhealthy measures of neighborhood food, physical activity, and social environments. *Prev Med* 2016; **93**: 21–6.

68 Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3·6 million adults in the UK. *Lancet Diabetes Endocrinol* 2018; **0**: 1–10.

69 The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med* 2017; 377:13-27.

70 Hu F. Obesity Epidemiology. Oxford: Oxford University Press, 2008.

71 Yusuf S, Rangarajan S, Teo K, *et al.* Cardiovascular risk and events in 17 low-, middle-, and high-income countries. *N Engl J Med* 2014; **371**: 818–27.

72 Fitzmaurice C, Allen C, Barber RM, *et al.* Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a Systematic Analysis for the Global Burden of Disease Study Global Burden . *JAMA Oncol* 2017; **3**: 524–48.

73 Secretan BL, Ph D, Scoccianti C, *et al.* Body fatness and cancer: viewpoint of the IARC working group. *N Engl J Med* 2016; **375**: 794–8.

74 World Health Organization. Urban green spaces and health: a review of the evidence. 2016.

75 Van Dyck D, Cerin E, De Bourdeaudhuij I, *et al.* Moderating effects of age, gender and education on the associations of perceived neighborhood environment attributes with accelerometer-based physical activity: the IPEN adult study. *Health Place* 2015; **36**: 65–73.

76 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *Br Med J* 2014; **348**: g1464.

77 Lovasi GS, Neckerman KM, Quinn JW, Weiss CC, Rundle A. Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index. *Am J Public Health* 2009; **99**: 279–84.

78 McCormack GR, Shiell A, Doyle-Baker PK, Friedenreich CM, Sandalack BA. Subpopulation differences in the association between neighborhood urban form and neighborhood-based physical activity. *Health Place* 2014; **28**: 109–15.

79 Forsyth A, Michael Oakes J, Lee B, Schmitz KH. The built environment, walking, and physical activity: is the environment more important to some people than others? *Transp Res Part D Transp Environ* 2009; **14**: 42–9.

80 Auchincloss AH, Diez Roux A V, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch Intern Med* 2009; **169**: 1698–704.

81 Thornton LE, Lamb KE, Tseng M, Crawford DA, Ball K. Does food store access modify associations between intrapersonal factors and fruit and vegetable consumption? *Eur J Clin Nutr* 2015; **69**: 902–6.

82 Inagami S, Cohen DA, Brown AF, Asch SM. Body mass index, neighborhood fast food and restaurant concentration, and car ownership. *J Urban Heal* 2009; **86**: 683–95.

83 Zenk SN, Schulz AJ, Izumi BT, Mentz G, Israel BA, Lockett M. Neighborhood food environment role in modifying psychosocial stress-diet relationships. *Appetite* 2013; **65**: 170–7.

84 Brown AF, Vargas RB, Ang A, Pebley AR. The neighborhood food resource environment and the health of residents with chronic conditions: the food resource environment and the health of residents. *J Gen Intern Med* 2008; **23**: 1137–44.

85 Ding D, Sallis JF, Conway TL, *et al.* Interactive effects of built environment and psychosocial attributes on physical activity: a test of ecological models. *Ann Behav Med* 2012; **44**: 365–74.

86 Stafford M, Cummins S, Macintyre S, Ellaway A, Marmot M. Gender differences in the associations between health and neighbourhood environment. *Soc Sci Med* 2005; **60**: 1681–92.

87 Stafford M, Marmot M. Neighbourhood deprivation and health: does it affect us all equally? *Int J Epidemiol* 2003; **32**: 357–66.

88 Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. *Int J Obes Relat Metab Disord* 2004; **28 Suppl 3**: S2-9.

89 Waalen J. The genetics of human obesity. *Transl Res* 2014; **164**: 293–301.

90 Richmond RC, Timpson NJ. Recent Findings on the Genetics of Obesity: is there public health relevance? *Curr Nutr Rep* 2012; **1**: 239–48.

91 Van Der Klaauw AA, Farooqi IS. The hunger genes: pathways to obesity. *Cell* 2015; **161**: 119–32.

92 Andreasen CH, Andersen G. Gene-environment interactions and obesity: further aspects of genomewide association studies. *Nutrition* 2009; **25**: 998–1003.

93 Boardman JD, Daw J, Freese J. Defining the environment in gene-environment research: lessons from social epidemiology. *Am J Public Health* 2013; **103**: S64-72.

94 Horn EE, Turkheimer E, Strachan E, Duncan GE. Behavioral and environmental modification of the genetic influence on body mass index: a twin study. *Behav Genet* 2015; **45**: 409–26.

95 Kowaleski-Jones L, Brown BB, Fan JX, Hanson HA, Smith KR, Zick CD. The joint effects of family risk of obesity and neighborhood environment on obesity among women. *Soc Sci Med* 2017; **195**: 17-24

96 Liu H, Guo G. Lifetime socioeconomic status, historical context, and genetic inheritance in shaping body mass in middle and late adulthood. *Am Sociol Rev* 2015; **80**: 705–37.

97 Mooney SJ, Grady ST, Sotoodehnia N, *et al.* In the wrong place with the wrong SNP: the association between stressful neighborhoods and cardiac arrest within beta-2-adrenergic receptor variants. *Epidemiology* 2016; **27**.

98 Owen G, Jones K, Harris R. Does neighbourhood deprivation affect the genetic influence on body mass? *Soc Sci Med* 2017. **185**:38-45

99 Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. *Proc Natl Acad Sci U S A* 2015; **112**: 354–9.

100 Davis OSP, Haworth CMA, Lewis CM, Plomin R. Visual analysis of geocoded twin data puts nature and nurture on the map. *Mol Psychiatry* 2012; **17**: 867–74.

101 Llewellyn C, Wardle J. Behavioral susceptibility to obesity: Gene-environment interplay in the development of weight. *Physiol Behav* 2015; **152**:494-501

102 Drewnowski A. Obesity, diets, and social inequalities. *Nutr Rev* 2009; **67**: S36–9.

103 Cannuscio CC, Hillier A, Karpyn A, Glanz K. The social dynamics of healthy food shopping and store choice in an urban environment. *Soc Sci Med* 2014; **122**: 13–20.

104 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev* 2011; **12**: e95–106.

105 Tseng M, Thornton LE, Lamb KE, Ball K, Crawford D. Is neighbourhood obesogenicity associated with body mass index in women? Application of an obesogenicity index in socioeconomically disadvantaged neighbourhoods. *Health Place* 2014; **30**: 20–7.

106 Frank LD, Saelens BE, Chapman J, *et al.* Objective assessment of obesogenic environments in youth: geographic information system methods and spatial findings from the Neighborhood Impact on Kids study. *Am J Prev Med* 2012; **42**: e47-55.

107 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Neighbourhood typologies and associations with body mass index and obesity: A cross-sectional study. *Prev Med* 2018; **111**: 351–7.

108 Meyer KA, Boone-Heinonen J, Duffey KJ, *et al.* Combined measure of neighborhood food and physical activity environments and weight-related outcomes: the CARDIA study. *Heal Place* 2015; **33**: 9–18.

109 Boone-Heinonen J, Diez-Roux A V, Goff DC, *et al.* The neighborhood energy balance equation: does neighborhood food retail environment plus physical activity environment = obesity? The CARDIA Study. *PLoS One* 2013; **8**: 12.

110 Hobbs M, Green MA, Griffiths C, *et al.* Access and quality of parks and associations with obesity: A cross-sectional study. *SSM - Popul Heal* 2017; **3**: 722–9.

111 Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation* 2012; **125**: 729–37.

112 Economos CD, Hatfield DP, King AC, Ayala GX, Ann Pentz M. Food and physical activity environments: an energy balance approach for research and practice. *Am J Prev Med* 2015; **48**: 620–9.

113 Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *Int J Behav Nutr Phys Act* 2007; **4**: 32.

114 Wang D, Brown G, Liu Y. The physical and non-physical factors that influence perceived access to urban parks. *Landsc Urban Plan* 2015; **133**: 53–66.

115 Barrington WE, Stafford M, Hamer M, Beresford SAA, Koepsell T, Steptoe A. Neighborhood socioeconomic deprivation, perceived neighborhood factors, and cortisol responses to induced stress among healthy adults. *Heal Place* 2014; **27**: 120–6.

116 Panter J, Jones A, Hillsdon M. Equity of access to physical activity facilities in an English city. *Prev Med* 2008; **46**: 303–7.

117 Ellaway A, Lamb KE, Ferguson NS, Ogilvie D. Associations between access to recreational physical activity facilities and body mass index in Scottish adults. *BMC Public Health* 2016; **16**: 756.

118 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Associations between the combined physical activity environment, socioeconomic status, and obesity: a cross-sectional study. *Perspect Public Health* 2018; **138**: 169–72.

119 Ding D, Adams MA, Sallis JF, *et al.* Perceived neighborhood environment and physical activity in 11 countries: do associations differ by country? *Int J Behav Nutr Phys Act* 2013; **10**: 1.

120 Burgoine T, Sarkar C, Webster CJ, Monsivais P. Examining the interaction of fast-food outlet exposure and income on diet and obesity: evidence from 51,361 UK Biobank participants. *Int J Behav Nutr Phys Act* 2018; **15**: 1–12.

121 Bodicoat DH, Carter P, Comber A, *et al.* Is the number of fast-food outlets in the neighbourhood related to screen-detected type 2 diabetes mellitus and associated risk factors? *Public Health Nutr* 2015; **18**: 1698–705.

122 Burgoine T, Forouhi NG, Griffin SJ, Brage S, Wareham NJ, Monsivais P. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study. *Am J Clin Nutr* 2016; **103**: 1540-1547.

123 Harrison F, Jones AP, van Sluijs EMF, Cassidy A, Bentham G, Griffin SJ. Environmental correlates of adiposity in 9-10 year old children: considering home and school neighbourhoods and routes to school. *Soc Sci Med* 2011; **72**: 1411–9.

124 Burgoine T, Alvanides S, Lake AA. Assessing the obesogenic environment of North East England. *Heal Place* 2011; **17**: 738–47.

125 Green MA, Radley D, Lomax N, Morris MA, Griffiths C. Is adolescent body mass index and waist circumference associated with the food environments surrounding schools and homes? A longitudinal analysis. *BMC Public Health* 2018; **18**: 482.

126 Griffiths C, Frearson A, Taylor A, Radley D, Cooke C. A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK. *Int J Behav Nutr Phys Act* 2014; **11**: 138.

127 Spence JC, Lee RE. Toward a comprehensive model of physical activity. *Psychol Sport Exerc* 2003; **4**: 7–24.

128 Van Tuyckom C. Macro-environmental factors associated with leisure-time physical activity: a cross-national analysis of EU countries. *Scand J Public Health* 2011; **39**: 419–26.

129 Wicker P, Downward P, Lera-López F. Does regional disadvantage affect health-related sport and physical activity level? A multi-level analysis of individual behaviour. *Eur J Sport Sci* 2017; **17**: 1350–9.

130 Kokolakakis T, Lera-Lopez F, Castellanos P. Regional differences in sports participation: the case of local authorities in England. *Int J Sport Financ* 2014; **9**: 149–71.

131 Laverty AA, Thompson H, Cetateanu A, Filippidis FT. Macro-environmental factors and physical activity in 28 European Union countries. *Eur J Public Health* 2018; **28**: 300–2.

132 Kokolakakis T, Castellanos-García P, Lera-López F. Differences in formal and informal sports participation at regional level in England. *Int J Sport Policy* 2017; **9**: 491–504.

133 Haase A, Steptoe A, Sallis JF, Wardle J. Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. *Prev Med* 2004; **39**: 182–90.

134 Coutts C, Chapin T, Horner M, Taylor C. County-level effects of green space access on physical activity. *J Phys Act Heal* 2013; **10**: 232–40.

135 Blok DJ, de Vlas SJ, van Empelen P, Richardus JH, van Lenthe FJ. Changes in smoking, sports participation and overweight: does neighborhood prevalence matter? *Heal Place* 2013; **23**: 33–8.

136 Carroll SJ, Niyonsenga T, Coffee NT, Taylor AW, Daniel M. Associations between local descriptive norms for overweight/obesity and insufficient fruit intake, individual-level diet, and 10-year change in body mass index and glycosylated haemoglobin in an Australian cohort. 2018; **15**: 1–16.

137 Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses: power and sample size for the interaction test. *J Clin Epidemiol* 2004; **57**: 229–36.

138 Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine: reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007; **357**: 2189–94.

139 Baker C. Obesity Statistics. *House of Commons Briefing Paper* 2018.

140 Allender S, Rayner M. The burden of overweight and obesity-related ill health in the UK. *Obes Rev* 2007; **8**: 467–73.

Chapter 2. DATA

In this chapter I introduce the UK Biobank resource, the principal source of data used throughout this thesis. I describe the establishment of the UK Biobank resource, the recruitment of participants and a summary of characteristics of the cohort, and I describe the data available to researchers, including various linked data sources that form the wider UK Biobank resource. I also provide details of additional sources of publicly available data that I linked to the UK Biobank cohort myself, for the analyses presented in Chapter 7. I then describe the process of data access and management that I undertook in preparation for data analysis. Details of the methods used to analyse the data, including details of the operationalisation of variables, are provided later, in Chapter 3 (Methods).

2.1. UK Biobank

2.1.1. Background

Understanding the determinants of common, complex health conditions requires the recognition that these conditions tend to be caused by a multitude of exposures that may each have modest effects and interact with each other in complex ways[1]. To investigate a wide range of exposures and their potentially complex interplay, large prospective studies with extensive, detailed data collection, broad distribution of exposures, and sufficient numbers of disease cases are needed[2].

To this end, the UK Biobank was established by the Medical Research Council and Wellcome Trust. UK Biobank is one of the largest and most detailed population-based, prospective cohorts ever established. It was designed to combine extensive, precise baseline assessment of exposures with comprehensive follow-up and characterisation of many different health-related outcomes in order to enable exploration of the many determinants of diseases of middle and older age[1]. The cohort comprises half a million adults aged 40-69 years at recruitment, and data collection spanned biological sampling, behavioural and life history data, and sociodemographic characteristics, along with ongoing prospective linkage to routine data collected through the National Health Service (NHS). This makes it a rich resource for investigating the determinants of a wide range of important diseases[1], and provides researchers with opportunities to investigate a wide range of disease outcomes through linkage to routine medical records that will increase over time[3]. The core cohort has also been linked to environmental data, principally a high-resolution spatial database of objectively measured characteristics of the physical environment around each participant's place of residence, referred to as the UK Biobank Urban Morphometric Platform (UKBUMP). Repeat assessment and additional data

collection has been carried out for subsamples of the cohort, and there is potential for future additional data collection and linkage to additional external data sources. To promote innovative science, UK Biobank has been made available to researchers globally[4].

In the following sections I describe in more detail the establishment of the UK Biobank resource, and the various data available to researchers, with a focus on the data I have used in this PhD project (see also Figure 2.2 at the end of this chapter).

2.1.2. Pilot phases

To test the feasibility of the proposed establishment of UK Biobank and its assessment methods, two piloting phases were undertaken[2]. The first was a small-scale pilot conducted on 300 participants in early 2005, to test key parameters of the baseline assessment protocol. The second pilot phase, conducted during March to June 2006, involved recruitment of 3,799 participants from the Stockport area to a single assessment centre. In this second pilot phase, all of the planned procedures were assessed, including: identification of the target sample from NHS records; invitation to participate; consenting and baseline assessment of potential participants; and collection, transfer and storage of data and biological samples. This pilot phase also allowed the determination of the response to invitation, and identification of any major factors that affected it, and assessment of participants' views on the baseline assessment visit, as well as an evaluation of their understanding of the consent to participate. This pilot study showed that protocols worked well, and provided information that was later used to refine the invitation and assessment procedures for the final protocol. Participants from the second pilot phase and their data were integrated into the final cohort.

2.1.3. Recruitment

2.1.3.1. Invitation and response

In the UK, 98% of the population are registered with a general practitioner (GP) through the NHS[4]. NHS records were used to identify 9.2 million people aged 40–69 years old who were living within 25 miles of one of the study's 22 assessment centres[1]. Between 2006 and 2010, an invitation letter with a provisional appointment to attend the nearest assessment centre was sent to each of these people, approximately 6–8 weeks ahead of the provisional appointment date. The invitation letter contained a detailed information leaflet describing the purpose of UK Biobank, how people had been identified for invitation, and what consenting to participate would involve. Invitees who wanted to find out more information were directed to a study website and free contact telephone number if they

wished to discuss the study with a member of the study team[2]. Overall, 503,325 participants were recruited, yielding a response fraction of 5.5%[1].

2.1.3.2. Assessment Centres

UK Biobank established 22 assessment centres in cities in England, Wales and Scotland, as displayed in Figure 2.1. Participants attended these centres to undergo the baseline assessment. The assessment centres covered a variety of settings, in order to maximise socioeconomic and ethnic heterogeneity, and ensure an urban–rural mix and broad distribution across all exposures. Assessment centres were necessarily located in cities, in order to be in close proximity to sufficient numbers of the target population. Assessment centres also needed to be conveniently located for public transport links, nearby parking and disabled access. The 22 assessment centres include the site of the integrated pilot phase (Stockport). Assessment was coordinated centrally and phased across the centres, with up to six assessment centres active at any one time during the recruitment phase (as shown in Table 2.1). Assessment centre staffing and equipment were configured to operate six days per week and assess approximately 100 participants per day at each centre[4].

Figure 2.1 Map of UK Biobank assessment centres

Table 2.1 Number of participants recruited to each UK Biobank assessment centre (sorted by dates of recruitment period)

Assessment centre	Recruitment dates	Participants*
Stockport (pilot)	13/03/2006 to 14/06/2006	3,799
Manchester	16/04/2007 to 22/12/2007	13,943
Oxford	30/04/2007 to 03/11/2007	14,063
Cardiff	08/10/2007 to 31/05/2008	17,885
Glasgow	16/07/2007 to 19/04/2008	18,653
Edinburgh	07/11/2007 to 07/06/2008	17,202
Stoke	05/12/2007 to 26/07/2008	19,441
Reading	14/05/2008 to 02/05/2009	29,426
Bury	14/01/2008 to 20/12/2008	28,326
Newcastle	23/01/2008 to 28/03/2009	37,011
Leeds	27/02/2008 to 11/07/2009	44,220
Bristol	09/07/2008 to 28/11/2009	43,020
Central London	27/08/2008 to 29/08/2009	12,584
Nottingham	30/07/2008 to 12/09/2009	33,883
Sheffield	05/08/2009 to 13/07/2010	30,399
Liverpool	28/01/2009 to 01/04/2010	32,825
Middlesbrough	29/04/2009 to 06/02/2010	21,290
Hounslow	17/06/2009 to 26/06/2010	28,881
Croydon	24/09/2009 to 09/07/2010	27,392
Birmingham	29/10/2009 to 21/07/2010	25,506
Swansea	11/03/2010 to 03/07/2010	2,284
Wrexham	16/08/2010 to 01/10/2010	649

* Totals as reported in UK Biobank Showcase. Excludes people initially recruited but not included in the final dataset released to researchers.

Source: http://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=UKB_centres_map (Accessed 7 May 2019)

2.1.4. Individual Assessment

The baseline assessment process of UK Biobank took between two and three hours per participant, and consisted of four parts:

- written consent;
- touch screen questionnaire;
- face-to-face interview;
- measurements.

Participants first provided electronic signed consent to the collection of baseline data and prospective linkage to administrative health records (see section 3.4.6 for details of UK Biobank's ethical approvals). Participants then proceeded to a touch screen questionnaire, which allowed them more privacy to answer questions, and to complete the form in their own time. It included questions about socio-demographics, lifestyle and behaviour, early life factors, psychosocial factors, general health and medical history, and also involved hearing and cognitive function tests. Assistance was available for participants when required.

Following the touch screen questionnaire, information collected in the questionnaire about serious illness, operations and other procedures, medications, employment and early life factors was then discussed in more detail with participants in a nurse-led face-to-face interview.

The following measurements were then taken:

- standing and sitting height;
- waist and hip circumference;
- weight and bioimpedance;
- hand grip;
- spirometry;
- bone density;
- arterial stiffness;
- eye measurements;
- fitness assessment using a static bike;
- detailed web-based diet questionnaire.

Each participant also donated blood and urine. Some participants also provided saliva samples. The biological samples were used to assay a panel of biomarkers, and the blood

samples were also used to perform genotyping assays for each participant, generating genome-wide genetic data. The protocols used by UK Biobank in the collection and processing of the genetic data are described more fully in section 2.1.7 of this chapter.

For this PhD project, of the data collected directly from participants at the baseline assessment, I have used anthropometric measurements (height, weight, waist circumference and bioimpedance measures), various items from the touch screen questionnaire, and genetic data. The release of biomarker data was delayed until 2019, so this was not available for use in this PhD. Subsamples of the main cohort have participated in detailed follow-up and enhanced phenotyping and exposure measurement[4], but those data have not been used in this thesis. Full details of the specific variables I have used or derived are provided in Chapter 3 (section 3.3).

2.1.5. The UK Biobank Urban Morphometric Platform (UKBUMP)

Linked to the individual-level UK Biobank dataset is a high-resolution spatial database of a wide range of objectively measured characteristics of the physical environment around each participant's place of residence. Scaling up a pilot study conducted in Wales, Sarkar and colleagues[5] automated and applied a series of spatial and network analyses to a number of UK-wide spatial databases in order to generate a collection of metrics characterising potentially health-promoting or health-damaging morphological features of the built environment surrounding the precise, geocoded home address of each individual in the UK Biobank. This collection of metrics is referred to as the UK Biobank Urban Morphometric Platform (UKBUMP), and full details are provided on the UK Biobank website (http://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=100115). The UKBUMP is the primary source of the neighbourhood data I use in this thesis.

Table 2.2 summarises the 'morphometrics' developed for the platform and the spatial databases from which these metrics were derived. The various UK-wide spatial databases from which the environmental metrics were derived include UK Ordnance Survey Mastermap and AddressBase Premium databases, UK Land registry data, digital terrain models and commercial satellite imagery, among others. Further details of the metric-generating processes used in the creation of UKBUMP can be found in Sarkar et al[5].

Table 2.2 Summary of data available in the UK Biobank Urban Morphometric Platform[5]

Morphometric category	Description	Measurement	Main source of raw data
Health specific destination accessibility	Proximity of home address to 39 potentially health-influencing destinations (including GP practices, dentists, fast-food outlets, industrial sites, schools, pubs, public transport, parks, among others	Street network distance in metres from home address to nearest destination in each category	Ordnance Survey AddressBase Premium
Land use density	Density of 217 different land use classes at 4 different levels of detail, for buffers around home address	Number of features of each land-use type per buffer, for buffers of 0.5, 1.0 and 2.0 km street network buffers, and for Lower Super Output Areas, around participants' home addresses*	Ordnance Survey AddressBase Premium
Accessibility foodscapes (London area only)	Proximity and density measures for 19 typologies of food stores	Street network distance to nearest address, and density in street-network buffers (0.5,1.0 and 1.5km)	Ordnance Survey AddressBase Premium
Greenness	Normalized Difference Vegetation Index (NDVI) of buffers around home address	Mean, minimum, maximum and standard deviation of NDVI for 0.5 and 1.0km Euclidean buffers around participants' home addresses	Bluesky International Ltd.
Street network accessibility	20 modelled indices of physical street-level accessibility for 19 different buffers around home address	Various	Ordnance Survey Mastermap Integrated Transport Network layer
Terrain (slope)	Terrain slope of area surrounding home address	Mean, minimum, maximum and standard deviation of slope(in degrees) for 0.5 and 1.0km Euclidean participants' home addresses	Bluesky International Ltd.
Building class	Description of the dwelling within which UK Biobank participant resides	Age, type, quality and class of building	The GeoInformation Group

* These data were released in two batches and one batch was provided as counts per km² rather than per buffer, so these were multiplied by the area of the buffer to produce a count.

In recognition of the challenges associated with delineating a health-relevant 'neighbourhood' for any given individual (discussed further in Chapter 3), the built environment was characterised in multiple ways and at multiple scales around each individual's home address. For land-use densities, a metric for each land-use type (e.g. residential, retail outlet, park, sports facility, manufacturing plant, public transport hub) was derived for street-network buffers of 500 m, 1000 m, 1500 m and 2000 m around each individual's place of residence, as well as the Lower Super Output Area (LSOA) in which the place of residence was located. Street-network buffers are catchment areas defined by distances along vehicle-access streets radiating from the home address, for example a 1000 m buffer will extend along each 1000 m stretch of street network from an individual's home address. In contrast, greenness and terrain metrics were derived for 500 m and 1000 m Euclidean buffers. Euclidean buffers are defined by straight-line (as-the-crow-flies) distances from the central address and are therefore circular, while street-network buffers are polygons defined by the specific road network of the area and influenced by physical boundaries such as waterways. While Euclidean buffers may be appropriate for measures to capture the overall greenness or terrain of an area, a street-network buffer is likely to be more suitable when considering local access to specific land-use types. Proximity metrics for health-specific destinations were defined not by buffers, but by distances along the street network from a person's home address to the nearest instance of each of 39 destination types, based on an origin-destination cost matrix algorithm to identify the 'least-cost' route.

Most measures in UKBUMP were available for participants recruited through all assessment centres except Stockport, the area in which the pilot study was conducted (n=3,799). Two important exceptions that influenced the choice of variables used in this PhD are the measures of a wider range of food outlet types (which are only provided for London), and the measure of local 'greenness', based on the Normalised Difference Vegetation Index (NDVI) derived from satellite imagery (which was absent or only partially observed for several assessment areas).

Details of the specific data I have used from UKBUMP, the variables I have derived from them to operationalise the neighbourhood measures central to this project, and for what analytical purpose each variable has been derived, are provided in Chapter 3. Briefly though, because my primary focus in the thesis is on obesity-related outcomes, I have focussed my choice of exposure variables by selecting one measure of neighbourhood environment relating to each side of the energy balance equation; in other words, one exposure relating to physical activity (number of formal physical activity facilities within

a 1000 m buffer) and one relating to diet (proximity to nearest takeaway/fast-food outlet). I also make use of other UKBUMP data to derive measures of neighbourhood-level confounders (residential density in a 1000 m buffer) and modifiers (parks and open/green spaces in a 1000 m buffer). In deciding on the specific data and measures to use, I made an *a priori* determination of the best available measures in UKBUMP (noting that these have their limitations, as I discuss later), drawing on the existing literature on neighbourhood environment measurement. These decisions are detailed further in Chapters 3 and 4.

Though there are substantial advantages to this dataset including its size and linkage to a wide-range of objectively measured health outcomes, there are also limitations. Relying on pre-defined, 'off-the-shelf' environmental measures such as those in UKBUMP, especially when based on primary data collected for non-research purposes can be problematic[6,7]. The documentation associated with UKBUMP provides insufficient detail to assess the accuracy and validity of the underlying databases. More detailed discussion of the limitations of the specific metrics used in this thesis can be found in the relevant chapters, and in a general discussion of limitations of the thesis in Chapter 9. Limitations aside, this is the only sample of this size, and for this population of particular importance, for which such a detailed set of person-centred environmental data is available alongside objective and wide-ranging health-related data.

2.1.6. Additional greenspace measures linked to UK Biobank

In addition to the measures contained in the UKBUMP, in 2018 a further set of environmental measures derived by Wheeler and colleagues[8] were linked to the UK Biobank cohort and made available to approved researchers, and I have accessed these to derive one of the exposure variables used in the thesis.

Amongst these extra environmental data are measures of residential green and blue space exposure, estimated for participants living in England (but not Scotland or Wales), using the Generalised Land Use Database (GLUD) for England 2005. This database provides data on land use distribution for 2001 Census Output Areas in England, across a range of land use categories. For each output area, the area of the GLUD categories 'greenspace', 'domestic gardens' and 'water' were calculated as a percentage of the total of all land use types, and these have been intersected with 300 m and 1000 m Euclidean buffers around UK Biobank participant home locations to allocate area-weighted means of each land use percentage coverage to each participant. The derived measures available to UK Biobank researchers are therefore percentage values for 'greenspace', 'domestic gardens' and 'water' within 300m and 1000 m buffers around each participant's home address. 'Greenspace' in

GLUD includes all public or private vegetated areas larger than 5 m^2 in area, with the exception of domestic gardens, which are classified separately. 'Water' refers to lakes, rivers, etc. As explained in the documentation for these measures[8], there is no established distance within which green spaces are thought to influence health, but there is some evidence for the relevance of a 300 m threshold[9,10], and this has also been taken up in policy recommendations in the UK[11]. A 1000 m buffer has also been used in some studies and data for this buffer size were also provided.

For this thesis, I have used these data to derive a variable characterising total neighbourhood greenspace (including gardens), in preference to the incomplete NDVI data available in UKBUMP.

2.1.7. Genotyping data

UK Biobank collected genome-wide genotype data on all participants using two purpose-designed genotyping arrays. Initially, a subset of 49,950 participants involved in a substudy were genotyped by Affymetrix (now part of ThermoFisher Scientific) using the UK BiLEVE Axiom Array. Following this, 438,427 participants were genotyped using the closely-related UK Biobank Axiom Array. Both arrays were purpose-designed specifically for the UK Biobank genotyping project and share 95% of marker content. The released genotype dataset combines results from both arrays, and includes a total of 805,426 markers[12]. In addition to these markers, genotypes at a further ~96 million loci were imputed.

The marker content of the arrays was chosen to capture genome-wide genetic variation, and includes many markers with known associations with, or with possible indicated roles in, disease, along with a range of rare coding variants (minor allele frequency (MAF) <1%), and markers that provide good genome-wide coverage for imputation of common (MAF >5%) and low frequency (1-5% MAF) genotypes in European populations[12] that were not directly assayed.

Blood samples collected at the baseline assessment visit were stored at the UK Biobank facility in Stockport, UK, until DNA extraction and genotyping were commenced in November 2013. A comprehensive quality control process was designed specifically to accommodate the complex processing of large-scale dataset. The data released to approved researchers contains genotypes of 488,377 UK Biobank participants. Genotypes are missing from the remaining 3% of the sample because insufficient DNA was extracted from participants' blood samples.

While the majority (94%) of the UK Biobank cohort reported their ethnic background as "White" and most of those White British, ~22,000 individuals had a self-reported ethnic background originating outside Europe, and this creates strong population structure at the genetic level (because of differences in allele frequencies across populations with different ancestral backgrounds). This population structure requires epidemiological studies using the data to account for the ancestral background of participants. To this end, a principal component analysis (PCA) was performed by UK Biobank on a subset of 407,219 unrelated, high-quality samples, following which all samples were projected onto the principal components (PC) and their loadings, thus forming a set of PC scores for all samples in the cohort[12].

The PCs separate participants by axes of ancestral background, with each additional PC capturing population structure at progressively finer sub-continental geographic scales. Consistent with this, individuals with similar PC scores have similar self-reported ethnic backgrounds. These PCs are made available to researchers to include in analyses to control for the population substructure. Bycroft and colleagues[12] note that researchers may want to analyse only the subset of participants with White British ancestry to further reduce the risk of confounding due to differences in ancestral background. They also note that even the White British ancestry subset may still contain subtle structure present at sub-national scales, which the methods applied to UK Biobank were not able to detect. Further, Haworth and colleagues[13] found evidence that there remained geographic structure in genetic data that could not be accounted for using routine adjustment for assessment centre and the PCs. I examine this possible source of bias in Chapter 5.

I describe in Chapter 3 (Methods) the procedures for accessing and managing selected genetic data for numerous BMI-related variants of interest, which I make use of in the analyses in Chapter 5.

2.1.8. Prospective data linkage

At recruitment, participant consent included consent to the ongoing linkage of their baseline data, and any subsequent data collected, to administrative health records, including death registrations (Office for National Statistics), cancer registrations (national cancer registries), Hospital Episode Statistics (HES) (including admissions, diagnoses, procedures) and primary care records via the NHS. This data linkage is updated regularly and made available to UK Biobank researchers. In this thesis I make use of the HES data, and these were available up to January 2016 at the time I accessed them. Hospital

admissions are coded using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)[14].

2.1.9. Response & representativeness

UK Biobank achieved a baseline response fraction of 5.5%, after inviting over 9 million eligible people to participate[1]. This is lower than the 10% response achieved in the pilot phase and which the study team expected to match or exceed in the main recruitment phase[2]. It is not clear what the reasons are for this.

A comparison of participants and non-participants (those who were invited but did not participate) found that participants were more likely than non-participants to be female, older, and live in less socioeconomically deprived areas. Regional differences were also observed. In a comparison of sociodemographic, physical, lifestyle, and health-related characteristics of the cohort with summary statistics for the general population from the UK Census and nationally representative population-based surveys, participants in the cohort had fewer self-reported health conditions, were less likely to be obese, and were less likely to smoke or to drink alcohol on a daily basis. During early follow-up, mortality and incident cases of disease were generally lower in the UK Biobank cohort than in the general population[15]. In other words, there is evidence of a "healthy volunteer" selection bias in UK Biobank, and it is not representative of the general population, at least on some measures. The implications of this have been debated. While UK Biobank acknowledge that the cohort (like many large prospective cohorts) is not suitable for estimating generalisable prevalence or incidence rates of disease, they argue that exposure-outcome associations estimated from the cohort should be broadly generalisable[15]. Others have disputed this claim, citing the risk of collider bias if the probability of participation in the study is influenced by the exposure and the outcome[16].

2.1.10. Withdrawals and exclusions

UK Biobank participants are free to withdraw consent at any time, and researchers working with the data are instructed periodically to remove recently withdrawn participants from unpublished analyses. Between the time the data for the project associated with this PhD were first made available to me, and the submission of this thesis, 96 participants have withdrawn. Some withdrew after the paper in Chapter 4 was published, but prior to the submission of this thesis. Therefore, the number of participants reported to be available in the cohort as a whole is slightly higher in that paper compared with elsewhere in the thesis.

The 3,799 cohort members recruited during the integrated pilot phase are also excluded from all analyses in this thesis because UKBUMP environmental data were not available for these individuals. I also excluded nine participants aged less than 40 or more than 70 at the baseline assessment. While recruitment eligibility was restricted to people aged 40-69 years, by the time of assessment some people were aged 70, and they were not excluded.

This leaves a total available sample of N=498,747. Final analyses excluded observations that were missing data on key variables in any given analysis, so were based on smaller and variable sample sizes. In Chapter 3 (Methods), where I report summary statistics for key variables, these are based on the largest available sample size.

2.1.11. Characteristics of the UK Biobank cohort

As already described, the UK Biobank participants were aged between 40 and 69 years when invited to participate, and members of the achieved sample were more likely than the target population to be female, older, live in less socioeconomically deprived areas, and own their home. The mean age of the sample is 56.5 years and 54.4% are female. All but 5% of the cohort is of White ethnicity, which is similar to the general population in the same age range in the 2001 Census (94.5% White) but somewhat higher than in the 2011 Census (91.3% White)[15]. Analyses in this thesis represent the subsamples of the cohort with complete data on the variables relevant for each analysis. This raises additional concerns about representativeness. Table 2.3 summarises the sample for key sociodemographic characteristics, showing the degree of missingness on key variables. Additionally, each of the research papers in the thesis contains a table summarising the characteristics of the sample relevant for that chapter. Several analyses presented in this thesis are further restricted to participants living in urban areas (in particular, Chapters 6 and 7). The ethnic diversity of the urban subsample (94.9% White) was very similar to that of the larger sample used in analyses that included both urban and non-urban participants (94.6% White). This suggests a possible mismatch with the target population because people of White ethnicity are less likely than other ethnic groups to live in urban areas (for the population as a whole, 89% of the urban population was classified as White in 2001, and 83% in 2011[17]). However for the age range of the target sample, ethnic diversity might be expected to be more similar in urban and non-urban areas than in younger age groups, as there is greater ethnic diversity amongst younger ages[18].

The distribution of each of the key exposures, outcomes and effect modifier variables is summarised in Chapter 3 where each variable is described in detail.

Table 2.3 Characteristics of available sample of UK Biobank participants (n=498,747*)

Characteristic			Range	% missing
Age**	Mean (SD)	56.5 (8.1)	40–70	0.0
Sex (female)	%	54.4		0.0
Ethnicity				0.5
White	%	94.6		
South Asian/South Asian British	%	1.6		
Black/Black British	%	1.6		
Chinese/other(non-South)Asian	%	0.7		
Mixed: White/Black	%	0.2		
Mixed: White/Asian	%	0.2		
Mixed - detail unknown	%	0.2		
Other	%	0.9		
Income				14.7
Less than 18,000	%	22.9		
18,000 to 30,999	%	25.4		
31,000 to 51,999	%	26.0		
52,000 to 100,000	%	20.3		
Greater than 100,000	%	5.4		
Education				1.3
College or University degree	%	32.7		
A levels/AS levels or equivalent	%	11.2		
O levels/GCSEs or equivalent	%	21.4		
CSEs or equivalent	%	5.5		
NVQ or HND or HNC or equivalent	%	6.7		
Other professional qualifications	%	5.2		
None of the above	%	17.3		
Employment status				0.6
Paid employment or self-employed	%	57.5		
Retired	%	33.5		
Unable to work	%	3.4		
Unemployed	%	1.7		
Home duties/carer/student/volunteer/other	%	4.1		
Area deprivation (Townsend index)	Median (IQR)	-2.1 (-3.6–0.5)		0.1
Urbanicity (home postcode classified as urban)	%	86.1		1.0

Note: Due to rounding error, some percentages sum to more than 100%

* Summary statistics for the full sample of participants from the 21 assessment areas for which UKBUMP environmental data are available (excludes Stockport, the site of the integrated pilot).

** 40-69 year olds were invited to participate in UK Biobank but the recruited sample was aged 37-73 (>99% aged 40-69). The analytical sample included people aged 70 at the time of assessment, but excluded nine individuals with complete data who were aged <40 or >70.

2.1.12. Data access and management

The preliminary phase of application for access to the UK Biobank data commenced in November 2015 and approval to proceed to the main application phase was granted in January 2016. A detailed project proposal and data request was submitted in February 2016, with approval granted in July 2016 (Project number 17380). The individual baseline data and an initial release of the UKBUMP data were made available in October 2016. The access procedure involved secure access via the UK Biobank Access Management System to download and extract main and supplementary datasets. Once extracted, these datasets were converted to Stata (.dta) format (StataCorp LP, College Station, TX, USA) to enable further data management and analysis.

Following extraction of the individual-level baseline assessment data and UKBUMP data, I undertook the following process to construct a master dataset from which all analytical samples were derived.

1. Labelling of all received variables (in their raw form).
2. Generation of descriptive statistics for all received variables for comparison with the reported summary statistics in UK Biobank's 'Data Showcase', to verify that the data extraction and conversion steps had produced the expected data.
3. Substantive review of key variables to familiarise myself with the data.
4. Review of values and missing data for all key variables to check for duplicates, implausible values, etc. UK Biobank data has already undergone extensive cleaning and quality assurance checking, so this step served mostly as a check, and another way to become more familiar with the data and anticipate any issues that may arise for analysis (e.g. distributional or data sparsity issues).
5. Derivation of variables for use in analysis, including recoding. While many variables were received in a format that was close to suitable for analysis, most required minor recoding (e.g. of missing values from a numeric code to system missing). Other key variables such as the environmental exposures and some covariates required categorisation from a continuous form, or collapsing to a smaller number of categories than the original variable. Details of the final derived variables are given in Chapter 3 (Methods).
6. Unnecessary variables were dropped and baseline assessment data were linked to the variables derived from the UKBUMP to form a single dataset.

In the initial data release the built environment data (UKBUMP) were incomplete, allowing only preliminary data management and analysis, but having established the data

management processes it was straightforward to update the master dataset when the updated data release became available in June 2017. The analyses in Chapters 4 and 6 were based solely on this master dataset. Genetic data, updated hospital admissions data, and the additional source of greenspace measures from UK Biobank, along with the open access Local Authority-level data, were not accessed until early 2018, so these were managed separately and linked to the master dataset at a later stage, in preparation for the analyses in Chapters 5, 7 and 8.

Genetic data were extracted and converted to custom tables by Dr Jody Phelan at LSHTM. The genetic data were downloaded from the European Genome-phenome Archive (EGA) in the form of bgen files for each individual chromosome. These were decrypted using the EGA download software, EgaDemoClient. The sample file linking the genotypes with subject IDs were downloaded using the ukbgene software. Plink2 (v2.0) was used to convert these files into bed/bim/fam files. All relevant SNP data (exact allele dosages for genotypes assayed directly, and imputed allele dosages otherwise) were extracted and converted into VCF format using plink (v1.9). The VCF files were merged using bcftools and parsed into a custom table format using in-house Python scripts. I separately accessed the principal components from the PCA for population structure directly through the UK Biobank Access Management System, in the same manner as other baseline data.

Hospital Episode Statistics (HES) data linked to the cohort were downloaded as a separate bulk file from the UK Biobank Showcase, and linked to the cohort using a provided key. Following coding guidance provided at an 'Introduction to HES' training course run by the Administrative Data Research Centre for England, I manipulated the raw HES data to aggregate hospital episodes into admissions (a single admission may be made up of multiple episodes, e.g. if a patient is transferred between consultants or departments), and then extracted the primary diagnosis attached to each admission, in order to identify relevant ICD-10 codes for the analysis in Chapter 8.

Details of the access and management of additional external data sources are provided below.

2.2. Additional Data Sources

The fourth objective of this PhD was to examine possible geographical heterogeneity in associations between the neighbourhood built environment and adiposity across England, and to explore whether any such heterogeneity might be explained by locally varying contextual factors. To achieve this, I assigned each UK Biobank participant to the Local Authority District in which their address was located, and then linked Local Authority-level data from three publicly available external datasets to the cohort. This enabled me to examine whether selected attributes of these larger administrative areas modified the association between the neighbourhood environment and BMI.

Due to privacy restrictions, the exact address coordinates of participants are not routinely made available to researchers; instead approximate coordinates (grid references rounded to the nearest 1 km) are available. Using these approximate coordinates I geocoded participants and assigned them to the LAD in which they reside, using QGIS v2.14[19].

During this process I identified 91 address points that were well outside the geographical scope of the UK Biobank study. I therefore excluded these from the analysis of geographical heterogeneity (presented in Chapter 7) as their locations were considered unreliable.

Once each UK Biobank participant was assigned to a Local Authority, it was then possible to link the UK Biobank dataset to external, Local Authority-level data. Being administrative units, Local Authority Districts are well described in publicly available datasets spanning multiple domains. This provided opportunities to consider a range of Local Authority-level variables that might be important effect modifiers (or confounders). After canvassing various available datasets, I identified suitable variables in the following datasets:

- The Land Cover Atlas of the UK[20]: a compilation of the relative proportions of land cover types for each Local Authority District, using a broad classification scheme and derived from Coordination of Information on the Environment (CORINE) Land Cover data from 2012[21].
- Neighbourhood Statistics – Local Authority Model-Based Estimates of Healthy Lifestyles Behaviours, 2003-05[22]: estimates of the prevalence of obesity, smoking, fruit and vegetable consumption and binge drinking among adults in each Local Authority, derived from the Health Survey for England and modelled with

additional information from other administrative data sources to produce accurate small area statistics for each Local Authority.

- Regional gross disposable household income, UK 1997 to 2016: annual estimates of gross disposable household income (GDHI) per capita for a range of geographical levels (NUTS1-3 and Local Authorities), produced by the Office for National Statistics[23].

Details of the variables extracted from each of these three datasets are provided in Chapters 3 and 7. Those variables were then linked to the master UK Biobank dataset by merging on the unique Local Authority District identifying code.

Figure 2.2 summarises visually the data used in this PhD project. Having described these data sources in this chapter, I turn in the next chapter to a description of the measures and methods I used in the analysis of these data.

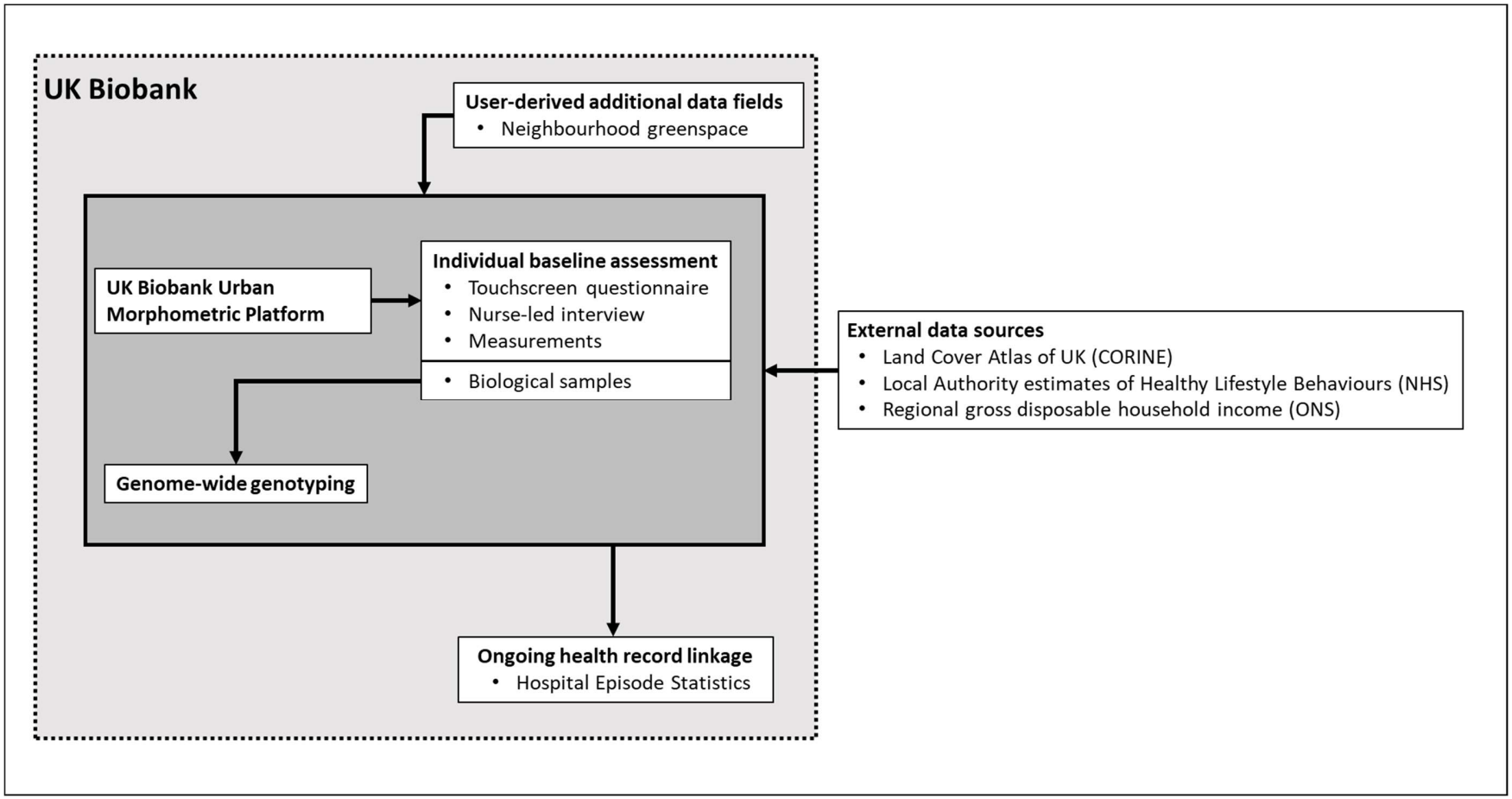


Figure 2.2 Visual summary of the UK Biobank resource and additional data sources used in this project

REFERENCES

1 Allen N, Sudlow C, Downey P, *et al.* UK Biobank: current status and what it means for epidemiology. *Heal Policy Technol* 2012; **1**: 123–6.

2 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. UK Biobank Coordinating Centre, 2007.

3 Collins R. What makes UK Biobank special? *Lancet* 2012; **379**: 1173–4.

4 Sudlow C, Gallacher J, Allen N, *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.

5 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP): a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21**: 135–48.

6 Cummins S, Macintyre S, Davidson S, Ellaway A. Measuring neighbourhood social and material context: generation and interpretation of ecological data from routine and non-routine sources. *Health Place* 2005; **11**: 249–60.

7 Wilkins EL, Morris MA, Radley D, Griffiths C. Using Geographic Information Systems to measure retail food environments: discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). *Health Place* 2017; **44**: 110–7.

8 Wheeler B. Environmental indicators attributed to participant's home location (UK Biobank Project 15374). 2016.

9 Grahn P, Stigsdotter UA. Landscape planning and stress. *Urban For Urban Green* 2003; **2**: 1–18.

10 Coles RW, Bussey SC. Urban forest landscapes in the UK: progressing the social agenda. *Landsc Urban Plan* 2000; **52**: 181–8.

11 Natural England. Nature Nearby – Accessible Natural Greenspace Guidance. 2010.

12 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**: 203–9.

13 Haworth S, Mitchell R, Corbin L, *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019; **10**: 333.

14 World Health Organization. International Classification of Diseases (ICD). 2011.

15 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**: 1026–1034.

16 Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2017; : 1–10.

17 Department for Environment Food & Rural Affairs. 2011 Census results for Rural England. 2013.

18 Office for National Statistics. England and Wales 2011 Census: UK population by ethnicity. 2018. Available from: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest

19 QGIS Development Team. QGIS Geographic Information System. 2016.

20 Rae A. A Land Cover Atlas of the UK. 2017. DOI:10.15131/shef.data.5266495.

21 Cole B, King S, Ogutu B, Palmer D, Smith G, Balzter H. Corine land cover 2012 for the UK, Jersey and Guernsey. 2015.

22 Scholes S, Pickering K, Rayat P, National Centre for Social Research. Neighbourhood Statistics, LA model-based estimates of healthy lifestyle behaviours - 2003-05. 2007. https://digital.nhs.uk/data-and-information/publications/statistical/neighbourhood-statistics-mental-health-adults-accessing-nhs-specialist-mental-health-services/neighbourhood-statistics-la-model-based-estimates-of-healthy-lifestyles-behaviours-2003-05 (accessed July 26, 2018).

23 Office for National Statistics. Regional gross disposable household income, UK 1997 to 2016. 2018.

Chapter 3. METHODS

3.1. Introduction

This thesis comprises five research papers using the UK Biobank data. Specific details of the methods employed in this thesis are provided in each paper, but due to the word limitations of the journals to which the papers are being submitted, the methods used are only briefly summarised in these papers. Therefore, in the current chapter I provide more detail on the methods used, including the definitions of key variables, how they were operationalised, and basic descriptive statistics.

Prior to that, I provide a brief expansion on some of the key concepts and definitions that have been used throughout this study, building on ideas introduced in Chapter 1. In particular, I discuss the concepts of effect modification and defining neighbourhoods in relation to health behaviours and outcomes.

3.2. Key definitions

3.2.1. Effect modification

Central to this thesis is the examination of effect modification in relation to the associations of interest between neighbourhood exposures and health outcomes. As stated in Chapter 1, when the strength and/or direction of an exposure-outcome relationship varies across values or strata of a third variable, that third variable is said to be an effect modifier, and we can infer that effect heterogeneity, or effect modification, is present[1].

Across the relevant literatures, various terms are used to describe the same methodological concept. The underlying idea that causal effects on an outcome might be heterogeneous across values of other variables is variously referred to as effect heterogeneity, effect modification, effect measure modification, moderation, and statistical interaction. Effect measure modification is technically most correct because for some classes of outcomes (e.g. binary) the choice of effect measure (e.g. rate ratio vs rate difference) can dictate whether effect modification is observed. Furthermore, from statistical models we might observe evidence that an effect estimate is modified, but if the model produces biased estimates of the true causal effect, then we cannot conclude with certainty than the true causal effect is in fact modified. Strictly speaking, conclusions are therefore drawn about whether an effect measure is modified, rather than an effect: hence, effect measure modification. In practice, effect modification is the more commonplace terminology, and in general, I use that term (and its derivatives) throughout this thesis when referring to specific relationships. I also use the more general 'effect heterogeneity'

when referring to the broader phenomenon. I avoid the use of the term moderation. While it is widely used and understood to be the same as modification in this context, it implies a tempering or dampening of a main effect, when in fact a third variable may either amplify or dampen the effects of the primary exposure.

As the testing of statistical interactions is a common way to assess effect modification, the term interaction is also often used interchangeably with any of the above-mentioned terms. Throughout this thesis I reserve the use of this term for instances where I am referring to interaction terms in regression models, or, in the case of Chapter 5 gene-environment interactions. It is important, however, to distinguish between statistical and biological interaction. In the latter case, the effect of the exposure on the outcome is biologically dependent on the presence or absence of a third variable, meaning both are necessarily causes of the outcome; in contrast, statistical interaction can occur whether or not the third variable (the modifier) is a cause of the outcome. Assessing questions of biological interaction therefore requires stronger causal assumptions and consideration of an additional adjustment set to minimise confounding[2]. In the thesis, when I use the term interaction it should be assumed to mean statistical interaction.

As stated, effect modifiers need not themselves be direct causes of the outcome. Though they may be, they might also act only by modifying the effect of another exposure. For example, faced with the same dose of a given exposure, men may be more likely than women to develop a given outcome, even though sex itself doesn't cause the outcome. In that case, sex would be considered an effect modifier of the exposure-outcome association. Alternatively, having a higher genetic risk of that outcome might make someone more sensitive to the same exposure than someone at lower genetic risk. In that case, genetic risk would be an effect modifier as well as a direct cause of the outcome. I have considered both kinds of potential effect modifiers.

3.2.2. Defining neighbourhood measures

As a convenient shorthand for referring to the local built environment around a person's home, I use the term 'neighbourhood' throughout this thesis. Depending on the specific measure being referred to, this takes on slightly different meanings, but in general it is used here to characterise exposure (or lack thereof) to health-promoting or health-damaging resources close to home.

Due to the reliance on secondary data in this thesis, the choice of environmental measures and the scales on which they are measured are all constrained by what is already available

within the UK Biobank resource, or what is able to be linked to it within the approved project scope. Working within the limits of the available data, this thesis is a collection of studies about relationships between health and particular characteristics of the built environment around cohort members' primary place of residence.

For the physical activity environment, I use a density measure of facilities based on a fixed buffer size around the home address. This kind of density-based definition is also employed for other environmental variables examined in this thesis as potential effect modifiers, or included as model covariates to control for potential confounding. The one notable exception is the measure of the fast-food environment that I employ in this thesis – proximity to a fast-food outlet. While I refer to this as a neighbourhood characteristic, it is not geographically bounded in the same way a density measure for a given buffer size is. Rather it is defined as the street-network distance from the home address to the nearest fast-food outlet, and, in some analyses, then categorised into levels of proximity. By categorising this proximity measure, it is also possible to indirectly interpret the results in terms of presence/absence of a fast-food outlet for the neighbourhood boundaries represented by the category cutpoints (500 m, 1000 m, etc.).

3.2.2.1. Challenges to defining health-relevant geographical areas

For any given measure, the ideal definition of a 'neighbourhood' is contested, and in reality it is likely that the geographical parameters of the health-relevant environment near home will vary from one individual to the next[3]. It may also vary for specific resources. For example, people might expect and be prepared to travel further to a formal physical activity facility such as a swimming pool, than to a local park. Thus, there is always a risk of what Duncan and colleagues[4] refer to as 'spatial misclassification', when a researcher arbitrarily defines a neighbourhood boundary in some way. Kwan[5] describes the Uncertain Geographic Context Problem (UGCoP), whereby the actual geographical area that exerts an influence on health is unknown for any given environmental characteristic, and furthermore the causally relevant timing and duration of exposure to that area are also unknown. For the physical activity environment, I use a street-network buffer size of one kilometre. Evidence suggests that one kilometre is the approximate distance from home that adults are typically willing to walk to reach places, and that areas within one kilometre of home are perceived by people to be part of their neighbourhood[6]. I use the same buffer size for neighbourhood availability of parks and other public green/open spaces (examined as an effect modifier in Chapter 6) and residential density (a model covariate). In the case of neighbourhood park availability, it was important to use the same buffer size as that used for formal PA facilities, because I was hypothesising that the two might, in a sense,

compete as health-promoting resources, with one moderating the influence of the other. In contrast, in Chapter 8 I use a 300 m buffer as a measure of neighbourhood exposure to greenspace, based on previous research showing this is a distance from home beyond which the use of green spaces quickly declines[7,8] and because it has been proposed in the UK as a benchmark for greenspace provision[9]. In that chapter, I was allowing that greenspace may act on cancer and CVD outcomes via pathways unrelated to physical activity, and was not directly relating the two environmental variables to one another in any way, so using a buffer with the stronger empirical basis for support was more important. When I capture the food environment using a proximity measure, no assumptions are made about the causally relevant geographical space with respect to fast-food exposure.

All studies that focus on the local residential area rest on an assumption that the residential neighbourhood, however defined, captures the relevant space in which people act and make decisions relating to behaviours such as diet and PA. It is increasingly being recognised that there are other important places – such as workplaces and commuting routes – that may also be relevant for many people, and that by ignoring these we may fall into the 'residential trap'[10]. This has given rise to the notion of 'activity spaces'[11], the measurement of which captures environmental exposures around additional anchor points beyond the home address, for example using travel surveys[12] or GPS tracking[13]. It is possible that the effects of exposures in the local residential neighbourhood – be they positive or negative – may be diluted by exposure to other areas, e.g. if your home neighbourhood lacks suitable places for physical activity, you might be able to compensate for this by making use of places near your workplace, such that characteristics of the workplace neighbourhood might modify or confound estimated effects of the home neighbourhood[14]. This poses challenges for constructing a well-defined exposure. Unfortunately it was not possible to account for other parts of individuals' activity spaces using UK Biobank.

Given that most of this thesis is focussed on investigating whether neighbourhood-health associations are modified by various individual and environmental modifiers, rather than estimating the magnitude of causal main effects of a neighbourhood exposure on a health outcome (with the exception of Chapter 4), some of these issues relating to the operationalisation of neighbourhood exposures may be less salient. Further discussion on the limitations of the particularities of neighbourhood definitions in relation to this project as a whole is included in the final chapter of the thesis.

3.3. Measures

3.3.1. Outcomes

Table 3.1 summarises the outcome measures used throughout the thesis, including their source and whether they are examined as cross-sectional or prospective outcomes.

Table 3.1 Outcome measures

OUTCOME	DETAILS
Body Mass Index (BMI)	Cross-sectional; measured at baseline assessment visit
Waist circumference (WC)	Cross-sectional; measured at baseline assessment visit
Percent Body Fat	Cross-sectional; measured at baseline assessment visit
Incident CVD-related hospital admissions	Prospective; identified in linked Hospital Episode Statistics data using ICD-10 codes I10-I25, I46,I48,I50,I60-79.
Incident cancer-related hospital admissions	Prospective; identified in linked Hospital Episode Statistics data using ICD-10 codes C00-C97 (excluding skin cancers (C43 and C44))

3.3.1.1. Adiposity

The primary outcome variables used in four of the five research papers in this thesis are measures of adiposity. Several objectively assessed adiposity measures are available in UK Biobank. Each has its own limitations and each performs differently in how well they predict various other health outcomes in various population subgroups[15], therefore I have examined three different measures rather than just one (when appropriate), and examined consistency of associations with the key exposures across these measures. The three adiposity measures examined in this thesis are waist circumference (WC), body mass index (BMI), and percent body fat.

3.3.1.2. Body Mass Index (BMI) and Percent Body Fat

Weight and various body composition measures were collected using a body composition analyser during baseline assessment[16]. This 'bioimpedance' equipment estimates body composition by measuring the resistance (impedance) to a small electrical current passed across body tissues between electrodes attached to the hands and feet, with greater impedance of the current indicative of more fatty tissue[16,17]. Impedance values are then used to automatically estimate percent body fat. BMI values (weight in kilograms, divided by square of height in metres) were calculated from the weight measurement taken by the

body composition analyser, and standing height collected using a calibrated height measure.

The bioimpedance machine was not used for participants who were pregnant, using a pacemaker, using a wheelchair, an amputee, unable to grip the handles of the machine, unable to stand, wearing a plaster cast or unwilling to remove their shoes. For these people, percent body fat is not recorded, and weight was instead measured by standard scales, from which BMI was then derived. Because some of the reasons for non-use of the bioimpedance machine would lead to inappropriate manually calculated BMI values, impedance-only BMI values were used in a sensitivity analysis in Chapter 4 to see if results were influenced by the method of BMI measurement.

BMI is a commonly used measure in epidemiological and clinical studies because it is easy to calculate. However, it is an indirect measure of body fatness because it does not distinguish fat mass from lean body mass. It therefore tends to be a poorer predictor of ill-health in older adults, and on average for a given BMI value, percent body fat tends to be higher in Asian populations than in White populations[17].

Limitations of the bioelectrical impedance method include that it is sensitive to the influence of hydration status, food intake and skin temperature. Percent body fat estimated by bioimpedance also tends to overestimate the true value in lean subjects while underestimating it in obese subjects[17].

3.3.1.3. Waist circumference

In contrast with BMI and percent body fat, which are both measures of overall adiposity, waist circumference (WC) is a measure of central or abdominal adiposity. Central adiposity is recognised as being more closely linked to cardiometabolic outcomes[18], and although WC can be more difficult to measure accurately, it is sometimes considered a more suitable measure than BMI for assessing risk of adiposity-related ill-health, particularly in older adults[17]. WC measurements were collected from participants using a tape measure and were manually recorded by the assessors.

All three adiposity metrics (BMI, WC and % body fat) have been treated as continuous variables in the analyses. I have elected to use continuous measures of adiposity rather than binary/categorical indicators of overweight or obesity, because the latter involves considerable loss of information[19]. Furthermore, while obesity is a strong risk factor for many health outcomes, the recent Global Burden of Disease study of the health effects of overweight and obesity estimated that more than a third of deaths and disability-adjusted

life years associated with a high BMI occurred in individuals who were overweight rather than obese, indicating the importance of considering the full range of BMI values, and other adiposity measures[20].

Figures 3.1–3.3 show the distribution of BMI, percent body fat and WC in the UK Biobank cohort. While the various adiposity measures are correlated (some more than others) (Table 3.2), a study in the UK recently found that patterns of social disparities in adiposity varied substantially depending on the measure of adiposity used[21], further highlighting the value of considering multiple measures.

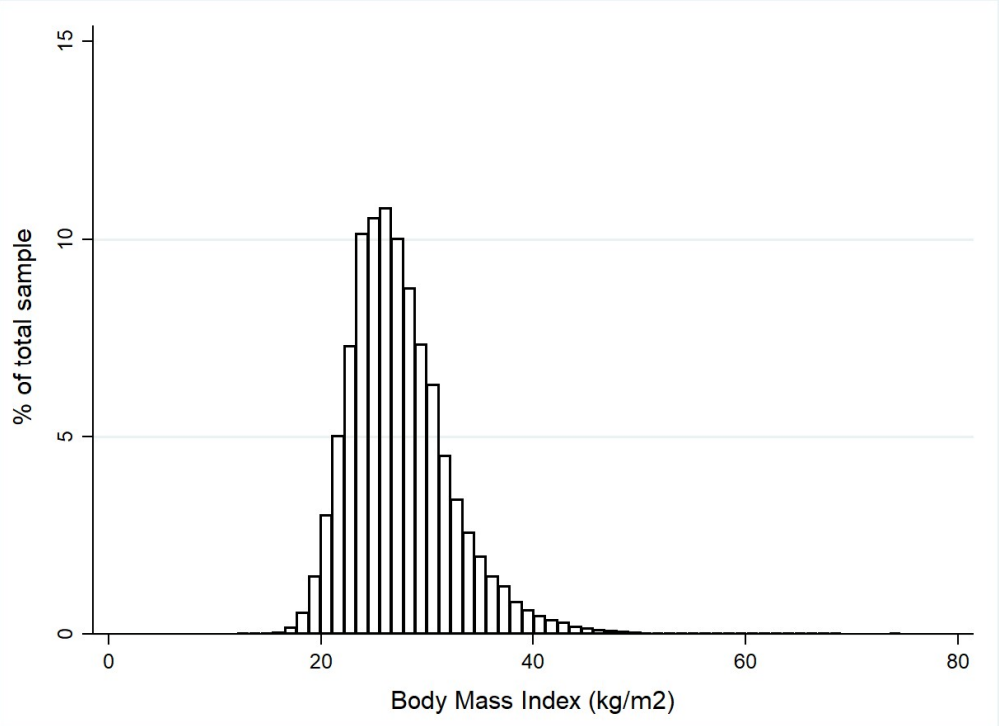


Figure 3.1 Distribution of BMI in the UK Biobank cohort

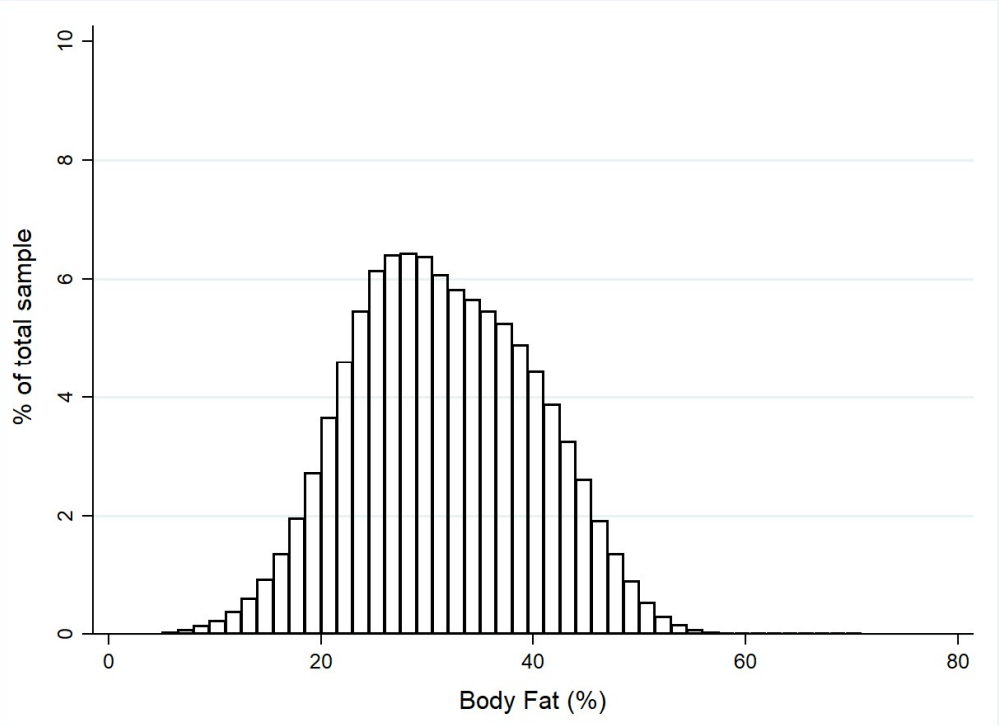


Figure 3.2 Distribution of percent body fat in the UK Biobank cohort

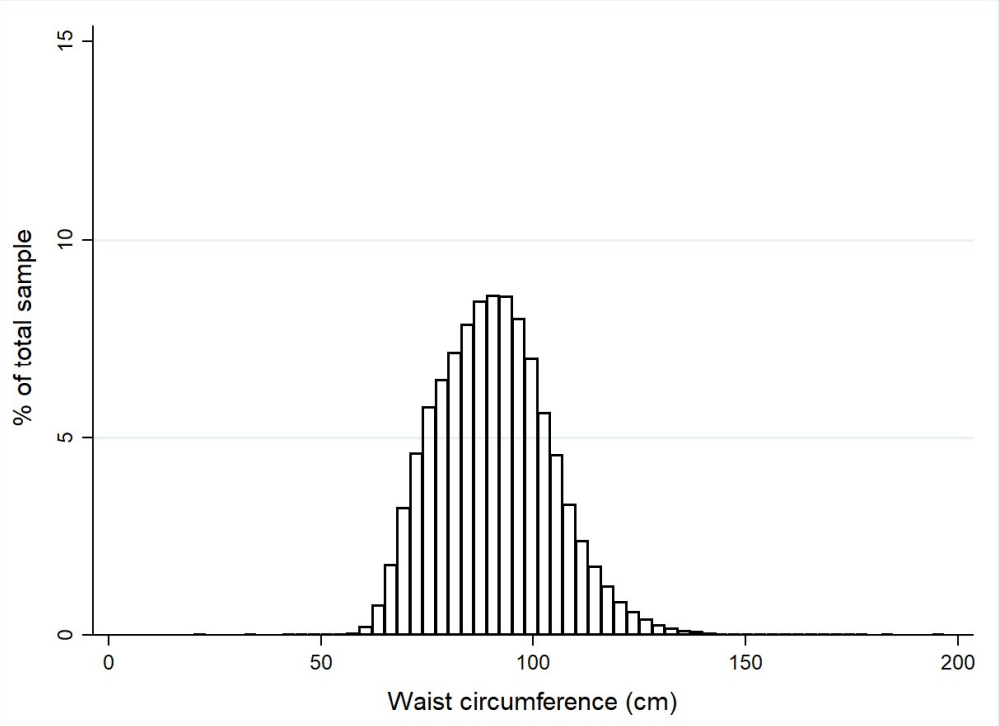


Figure 3.3 Distribution of waist circumference in the UK Biobank cohort

Table 3.2 Correlation matrix for BMI, WC and percent body fat in UK Biobank cohort

	BMI	WC	% Body Fat
BMI	1.0000		
WC	0.8137	1.0000	
% Body Fat	0.5715	0.2411	1.0000

One or all of these adiposity measures are the primary outcomes in the research papers in Chapters 4-7. I examine all three measures in Chapters 4 and 6. In Chapters 5 and 7 I focus on BMI. In Chapter 5 this is because I am examining gene-environment interactions with genetic variants and polygenic risk scores associated specifically with BMI. In Chapter 7 where I examine geographical heterogeneity of associations, I focus only on BMI, after seeing in Chapters 4 and 6 that results are consistent across the three adiposity measures.

The measures of BMI and percent body fat collected using the body composition analyser are missing in 2% of cases. For BMI, manual measurements are available for some of those missing the impedance measurements, reducing missingness on BMI to 0.6% of the total sample. Waist circumference is missing in 0.4% of the sample.

3.3.1.4. Hospital Admissions

In the final research paper (Chapter 8), I move away from cross-sectional associations with adiposity outcomes, and examine longitudinal associations with incident CVD- and cancer-related hospital admissions.

Hospital admissions were identified in the linked HES data, using the ICD-10 coding to identify cause of admission[22]. The specific outcomes were defined as any hospital admission for which the primary diagnosis is recorded as cardiovascular disease (ICD-10 codes I10-I25, I46, I48, I50, I60-79) or cancer (ICD-10 codes C00-C97, excluding skin cancers C43-44). CVD and cancer admissions were examined separately. I also separately examined associations with admission for two cancer subtypes that have the most well-established links to physical activity[23] and diet[24] – breast cancer (C19) and colorectal cancer (C18).

Table 3.3 Number of CVD- and cancer-related hospital admissions in UK Biobank cohort

	Number of admissions	Admissions as % of sample
CVD-related	26,984	5.4%
Cancer-related (excl. skin cancers)	26,495	5.3%

Note: Totals do not exclude individuals with pre-existing conditions

3.3.2. Environmental variables

In this thesis, I make use of several types of variables that capture characteristics of geographical areas (Table 3.4). Conceptually these represent two distinct geographical scales: the local residential environment, or 'neighbourhood'; and the wider context in which those more local residential neighbourhoods are located, which I refer to as the 'macro-environment'. Practically, these have needed to be operationalised based on available data, and this means that the neighbourhood variables are defined in slightly different ways. While all are defined with reference to the home address of an individual, they are a mix of land use density of a street-network buffer around the home address, percentage of a Euclidean buffer around the home address, street-network distance from the home address to a destination, or attributes of the postcode or census area in which the home address is located. The two macro-environmental variables are defined as attributes of the Local Authority District in which the home address is located, but unlike the neighbourhood variables, allocation to a Local Authority relied upon approximate rather than precise home location grid references. These variables are summarised in Table 3.4. Details of how each were defined are provided in the next section of this chapter according to their analytical purpose (exposure, effect modifier or confounder). Due to the reliance on secondary data, the timing of the collection of the underlying datasets varies. While I have made every attempt to use data as close as possible to the UK Biobank baseline phase, there remains some risk of temporal mismatch (discussed in Chapter 9).

One further geographical level at play in the UK Biobank is that of the assessment areas. Participants were sampled from within a 25-mile radius centred around 22 study centres across the UK where the baseline assessment took place. This clustered sampling design and the possible spatial dependence arising from it needs to be accounted for in analyses. To this end, assessment area is specified as the level-2 identifier in multilevel models in Chapters 4-7, also allowing the primary relationships of interest to vary by assessment area. In Chapter 8, dummy variables for assessment areas are instead included in single-level Cox proportional hazards models.

Table 3.4 Measures of environmental attributes

MEASURE	CONCEPTUAL GEOGRAPHICAL LEVEL	ANALYSED AS	DETAILS	DEFINED WITH REFERENCE TO	SOURCE
Formal physical activity environment	Neighbourhood environment	Primary exposure (Ch 4-8)	Land-use density measure of availability of facilities such as sports facilities, gyms, swimming pools (1000m street-network buffer)	Exact home address	UKBUMP Measured at end of baseline*
Fast-food environment	Neighbourhood environment	Primary exposure (Ch 4,5,7,8) Potential effect modifier (Ch 6)	Distance to nearest hot/cold takeaway/fast-food outlet	Exact home address	UKBUMP Measured at end of baseline*
Green space: Public greenspace and domestic gardens	Neighbourhood environment	Primary exposure (Ch 8) Secondary exposure (Ch 4)	Percentage of 300m buffers defined as 'public greenspace' or 'domestic garden' in GLUD	Exact home address	GLUD (2005)
Parks & public/open green spaces	Neighbourhood environment	Potential effect modifier (Ch 6)	Land-use density measures for parks and other public green/open spaces.	Exact home address	UKBUMP Measured at end of baseline*
Area deprivation	Neighbourhood environment	Potential confounder (Ch 4-8) Potential effect modifier (Ch 8)		Postcode of home address	Townsend Index (2001)
'Natural' land cover	Macro-environment	Potential effect modifier (Ch 7)	Percent of Local Authority District land cover classified as 'natural'	Approximate home address (rounded to nearest km)	Land Cover Atlas of the UK (based on CORINE) (2012)
Urbanicity	Neighbourhood environment	Potential confounder (Ch 4,5,8)	Home postcodes used to classify participants from England and Wales according to an ONS-defined scale of urbanicity collapsible to 4 categories based on population density (2001 Census). Scottish participants are classified using a scale that is not strictly equivalent, so I collapse all to urban/non-urban	Postcode of home address	UK Biobank (ONS) (population density based on 2001 Census)
Residential density	Neighbourhood environment	Potential confounder (Ch 4-8)	1000m street-network buffer	Exact home address	UKBUMP Measured at end of baseline*
Local descriptive obesity norm	Macro-environment	Potential effect modifier (Ch 7)	LAD adult obesity prevalence (published modelled estimates based on 2003-05 Health Survey for England data).	Approximate home address (rounded to nearest km)	Health Survey for England (2003-05)
LAD-income	Macro-environment	Potential confounder (Ch 7)	Local Authority Gross Domestic Household Income	Approximate home address (rounded to nearest km)	ONS (2006)

* Source material for UK Biobank Urban Morphometric Platform states the spatial data "were collected as close as possible to the end of the baseline wave to avoid temporal mismatch" (Sarkar et al. Annals of GIS, 2015, p.12) but does not provide specific dates, and cites the 2012 release of OS AddressBase Premium. Other publications by the creators of UKBUMP and using exposures derived from OS AddressBase state these "were assessed towards the end of the baseline phase (2010)" (Sarkar et al. Lancet Planetary Health, 2017, p.e279).

3.3.3. Exposures

Four of the five objectives of this thesis are focused exclusively on obesity-related outcomes, and as mentioned in Chapter 2, to address these objectives I focussed on two exposure variables, each thought to influence these outcomes via a separate pathway, and each relating to opposite sides of the energy balance equation: availability of formal PA facilities, and proximity to fast-food outlets. Availability of formal PA facilities near home is relatively understudied in relation to obesity-related outcomes, but what research there is suggests a possible relationship, so further research in this large UK-wide sample was warranted. Other features of the built environment are also thought to promote PA (e.g. walkability) but as these are the focus of extensive research by others, I have chosen to focus on the formal PA environment. Community food environments, including a focus on fast food, have received considerable research attention, but the evidence remains inconclusive. Of the limited data on the food environment available in UKBUMP for the full UK Biobank cohort, fast-food proximity was arguably the most appropriate for my analyses. Fast-food outlets and formal PA facilities are neighbourhood resources that are also both primarily commercial in nature, and therefore amenable to regulatory and market-based interventions. To address my fifth research objective, where my outcomes of interest were CVD and cancer, I also examined a third neighbourhood exposure (domestic and public greenspace) that might also influence these outcomes through pathways independent of adiposity or energy balance.

Availability of formal physical activity facilities is examined as an exposure in all five research papers. Proximity to a fast-food outlet is examined as an exposure in four research papers (Chapter 4, 5, 7 and 8), and also as a potential neighbourhood-level effect modifier in Chapter 6. Neighbourhood greenspace is examined as an exposure in the final research paper (Chapter 8). For completeness, I also examine this greenspace measure as a secondary exposure in additional material in Chapter 4, as it was not available at the time of publishing the research paper in that chapter. The following section explains the operationalisation of these three neighbourhood exposures in greater detail.

3.3.3.1. Availability of formal physical activity facilities

There is some evidence that neighbourhood exposure to dedicated facilities for physical activity, sometimes referred to as recreation facilities, is associated with higher levels of physical activity[25–27]. Whether they are associated with body weight outcomes is less well studied and less clear.

Definitions of such facilities usually include swimming pools, gyms, sports playing fields, leisure centres, and racquet sports facilities, among other facilities. These facilities may be open to the public or restricted to membership holders, and may be free at point of use or pay-to-use. Across studies, the definitions used usually varies somewhat, and there is some indication that the types of facility matter[28].

In this thesis, based on land-use density data from the UKBUMP, I define neighbourhood availability of formal physical activity facilities as the density (count) of physical activity facilities within a 1000 m street-network buffer around the home address. Physical activity facilities were defined as any land use classified in the Commercial-Leisure subcategory (CL06) of the UK Ordnance Survey AddressBase database. This subcategory comprises any address point classified as "Indoor/Outdoor Leisure/Sporting Activity/Centre not further defined", as well as the following more specific categories of land use:

- Bowls Facility
- Cricket Facility
- Diving / Swimming Facility
- Equestrian Sports Facility
- Football Facility
- Golf Facility
- Activity / Leisure / Sports Centre
- Playing Field
- Racquet Sports Facility
- Rugby Facility
- Recreation Ground
- Skateboarding Facility
- Civilian Firing Facility
- Tenpin Bowling Facility
- Water Sports Facility
- Winter Sports Facility

Non-commercial resources that may be considered informal physical activity facilities, such as public parks and other public open spaces, and walking and cycling paths, were not included in the count (except where covered by the above classification e.g. playing fields).

Collapsing all these facility types into a single measure may obscure facility-specific associations with health outcomes, but due to the large proportion of address points classified in the 'Indoor/Outdoor Leisure/Sporting Activity/Centre not further defined' category, it was decided that such a disaggregation would not be meaningful. While the land-use density is the only measure of PA facilities available in UK Biobank, the heterogeneous nature of this exposure makes a density measure more suitable than a proximity measure in this case.

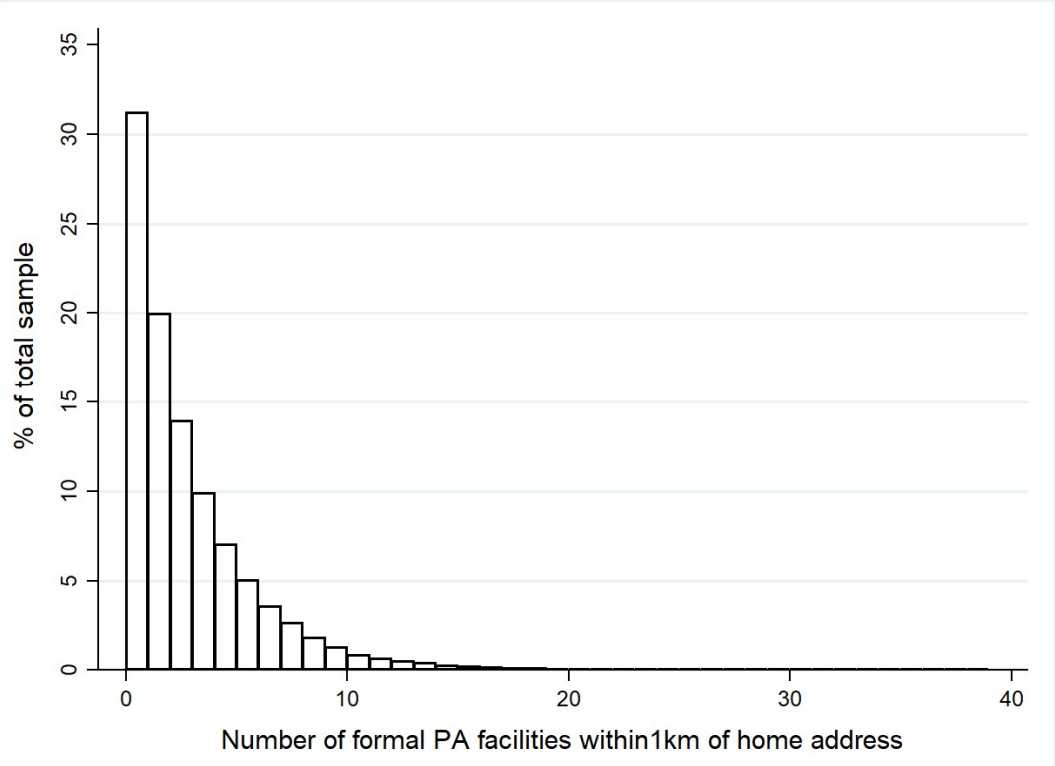


Figure 3.4 Distribution of availability of formal PA facilities within 1000m buffers

Data on the availability of PA facilities were missing for 3.5% of the full sample. For those with data, just under a third of participants had no formal physical activity facilities within a kilometre of their home. The median number of facilities within a neighbourhood buffer was one (IQR: 0-3). The full range was zero to 39, but 90% of participants had no more than six facilities near home (Figure 3.4).

In Chapter 4 this exposure variable was categorised as no facilities, 1 facility, 2-3 facilities, 4-5 facilities, or 6 or more facilities (Table 3.5). In Chapters 6 and 8, where the sample was further stratified by an effect modifier, the top two categories were collapsed so the uppermost category was '4 or more facilities', due to the number of observations in the '6 or more' category becoming quite small when stratified by another variable.

Table 3.5 Categories of neighbourhood availability of formal PA facilities

Number of formal PA facilities within 1km	n	%
0	150,187	30.11
1	96,018	19.25
2 - 3	114,677	22.99
4 - 5	58,212	11.67
6 or more	61,973	12.43
missing	17,680	3.54
Total	498,747	100.00

In Chapter 5 and Chapter 7 it was desirable to work with a continuous exposure for reasons specific to those chapters (outlined therein). In those two chapters, availability of PA facilities is therefore treated as a continuous variable, consistent with the roughly linear association observed using the categorical variable across the sample as a whole in Chapter 4. To treat it as a continuous variable, the number of facilities per 1000 m buffer has been topcoded due to the very long tail of the positively skewed distribution (0.07% of the total sample were recoded to 15 from a higher value; recoded distribution shown in Figure 3.5).

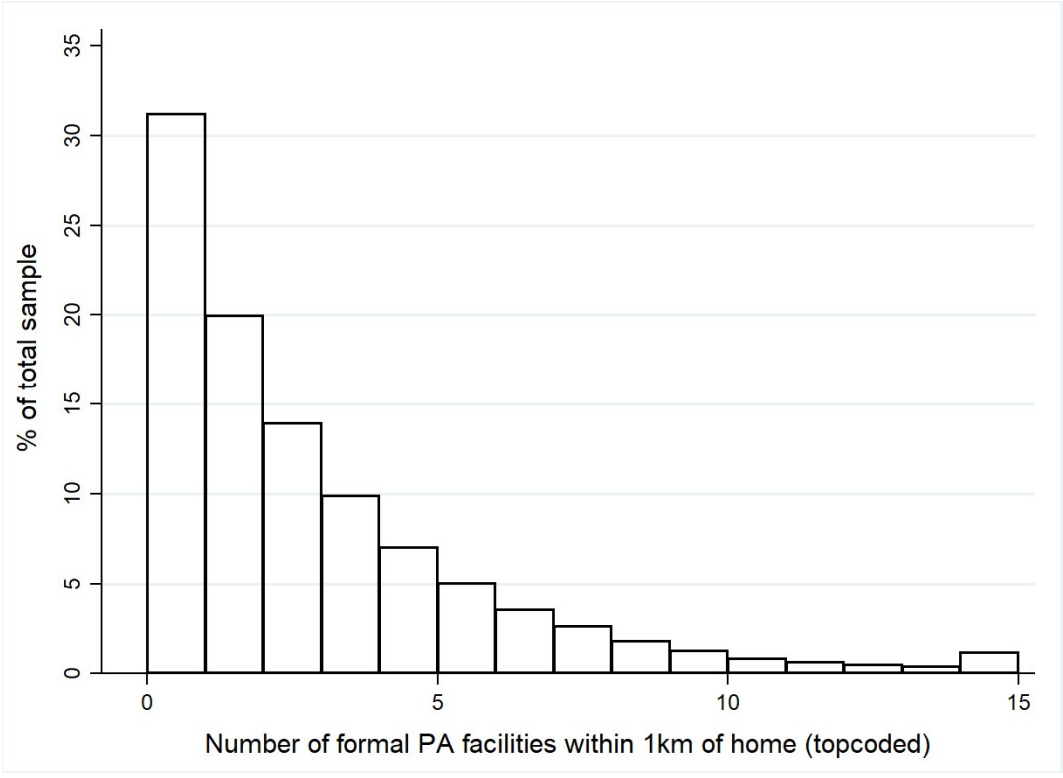


Figure 3.5 Topcoded distribution of availability of formal PA facilities within 1000m buffers

3.3.3.2. Fast-food environment

There is a large body of research investigating whether access to fast-food outlets near the home may be a determinant of weight status and obesity[29], but the conclusions have been inconsistent or equivocal, and most evidence comes from the USA, where the structure of the built environment differs from countries such as the UK. While a recent study in the UK found that exposure to fast-food outlets near home was associated with both BMI and odds of obesity in adults[30] this study was limited to one area in the UK and other international research has not consistently replicated findings from the USA[31,32].

Neighbourhood exposure to fast-food outlets can and has been operationalised in a variety of ways in different studies, including proximity measures (how close a person lives to a fast-food outlet), absolute count or density measures (number of outlets in a buffer, or outlets per population or km^2), and relative measures (e.g. fast-food outlets as a proportion of all food retailers in an area)[33]. Each captures something different about the fast-food environment that may plausibly influence diet and subsequent health outcomes.

In this thesis, exposure to fast-food outlets is operationalised as a proximity measure. Ideally, a relative measure of the food environment would have been used, but it was not possible to construct such a measure from the UKBUMP data for the full sample. Proximity has been used in numerous studies in the UK and elsewhere to capture exposure to fast-food outlets[31,34–36]. Proximity and density measures of the food environment have been shown to be moderately to strongly correlated in the UK[37] and while neither is demonstrably superior to the other, a recent systematic review (focussed on dietary outcomes) found that proximity measures of the food environment tend to produce smaller effect sizes than absolute density measures[38], indicating proximity measures may be a more conservative approach. Proximity measures also avoid the need to make assumptions about the causally-relevant geographical area. I therefore used the UKBUMP data on the distance from each participant's home location to the nearest address point in the UK Ordnance Survey AddressBase Premium database that was classified as a 'hot/cold fast-food outlet/takeaway'. Thus, for each individual, fast-food proximity is defined as the street-network distance (in metres) from their home address to the nearest takeaway/fast-food outlet. Data were missing for 3.6% of the full sample, including 20 observations with a distance greater than 40km to the nearest fast-food outlet, which I excluded on the grounds that these were implausible values given the catchment area for the study was itself roughly that distance around centrally located assessment centres.

Based roughly on the distribution of the data (Figure 3.6), I used these distances to categorise individuals as living <500 m, 500-999 m, 1000-1999 m, or 2000m+ from their nearest fast-food outlet (Table 3.6). This categorical form of the exposure was used in Chapters 4 and 8. As with the PA environment, in Chapter 5 and 7, I treated proximity to fast-food as a continuous variable, to better facilitate the analyses conducted in those chapters. When treating fast-food proximity as continuous, I used its logarithm (base 10) transformation for ease of interpretation, so that a one-unit increase represented a 10-fold increase in distance to the nearest outlet (e.g. 100m to 1000m) (Figure 3.7). To facilitate this I recoded to one metre a small number of observations where the distance to the nearest outlet was zero or less than one metre, because it is not possible to take the logarithm of zero, and the logarithms of values less than one are negative.

As a large distance to a fast-food outlet indicates that there are none close to the home address, while a smaller distance indicates there is at last one near the home address, this could alternatively be reclassified into a binary variable with a cutoff at e.g. 1km. Instead, by splitting distance into multiple categories, I make greater use of the available information while still allowing indirect interpretation in terms of presence/absence for the neighbourhood boundaries represented by the category cutpoints. This avoids making assumptions about the causally-relevant geographical area.

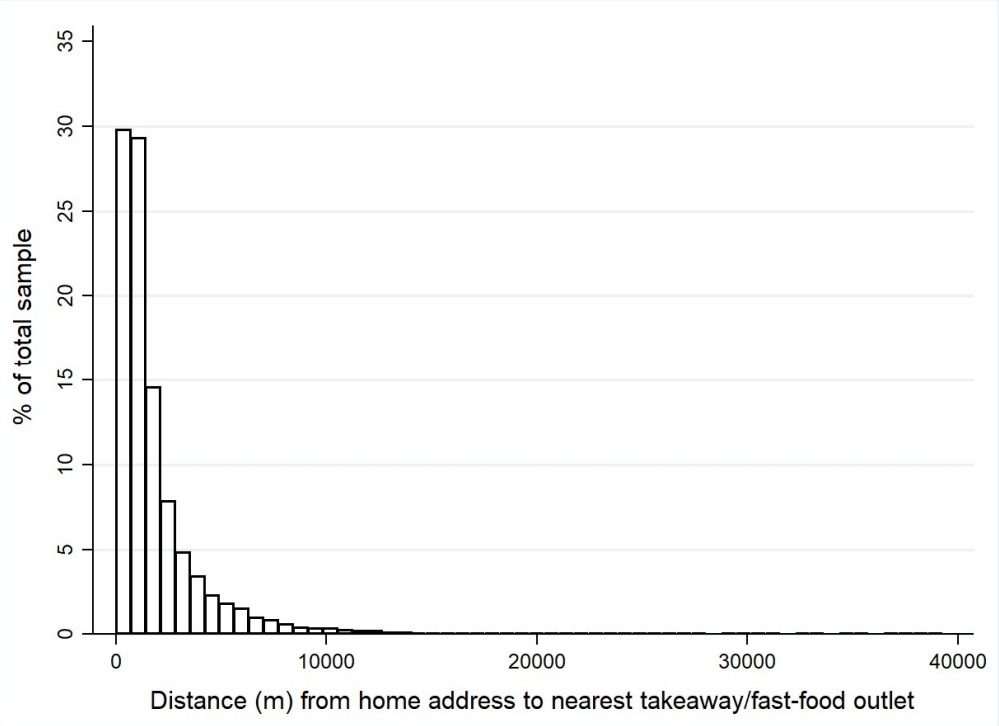


Figure 3.6 Distribution of proximity to nearest fast-food outlet

Table 3.6 Categories of proximity to nearest fast-food/takeaway outlet

Distance to nearest fast-food/takeaway outlet	n	%
Closer than 500m	88,789	17.80
500-999m	124,684	25.00
1000-1999m	133,386	26.74
At least 2000m	134,161	26.90
missing	17,727	3.55
Total	498,747	100.00

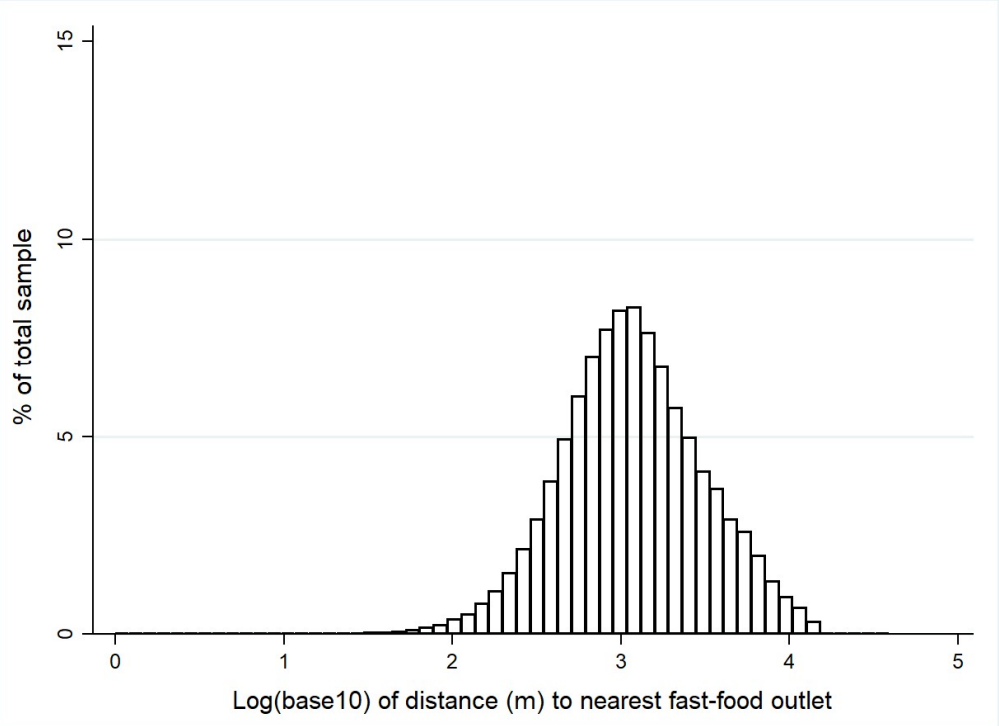


Figure 3.7 Logarithmic transformation of proximity to nearest fast-food outlet

Food outlet classification in the source database for the current study is supplied by local authorities and may include misclassification of some outlets (particularly misclassification of addresses as restaurants rather than fast-food outlets), potentially biasing the regression coefficients estimated in this thesis towards the null. The quality of the underlying source data – in terms of accuracy and completeness – could also vary geographically, introducing more misclassification error in some areas than in others. Unfortunately is was not possible to assess the accuracy and validity of the measure and its source data. I discuss the implications of this in the limitations section of the Discussion, and in individual research papers.

3.3.3.3. Greenspace

Unlike formal PA facilities and the fast-food environment, which will predominantly influence energy balance-related outcomes if they influence health at all, exposure to green spaces has considerable potential to influence health through multiple causal pathways. Various conceptual models linking green space to health have been proposed[39–41] and these typically recognise multiple pathways by which exposure to green space or nature more generally might influence health, including improved air quality, opportunities for physical activity, stress reduction and relaxation, resilience to heat-related illness by mitigation of the urban heat island effect, greater social cohesion, buffering from noise pollution, exposure to natural light, and improved functioning of the immune system[42].

Therefore, in this thesis, I focus on PA facilities and the fast-food environment when I consider obesity-related outcomes (Chapters 4-7), but in Chapter 8 when I consider a wider set of health outcomes (namely CVD and cancer), I also examine greenspace as an additional neighbourhood exposure because of its potential to influence these outcomes through multiple pathways.

A review[43] of recent studies of green space and health identified three dominant ways of objectively quantifying exposure to green space or greenness: vegetation indices based on satellite imagery (e.g. Normalised Difference Vegetation Index (NDVI)); metrics based on land-use databases (e.g. proportion of an area classified as some form of greenspace); and proximity metrics (e.g. the distance from a person's home to the nearest park). As with other neighbourhood exposures, each captures a slightly different dimension of what it might mean to be 'exposed' to 'greenspace'.

When examining greenspace as an exposure here, I make use of the additional measures added to UK Biobank derived from the Generalised Land Use Database (GLUD). I combined the land use percentages for 'greenspace' and 'domestic gardens' to derive a combined measure of greenspace exposure for 300 m neighbourhood buffers. Combining 'greenspace' and 'domestic gardens' is consistent with previous research using the GLUD to examine relationships with health[44]. Of the 300 m and 1000 m buffers available to UK Biobank researchers, I chose to use the 300 m buffer as the primary delineation of neighbourhood exposure to greenspace on the basis that previous research has shown this to be a distance beyond which the use of green spaces declines[7,8] and also because it is a policy-relevant distance, having been proposed in the UK as a benchmark for greenspace provision[9]. Furthermore, with the inclusion of gardens in the measure, a buffer of 300 m

is more likely to capture meaningful exposure to domestic gardens, as within a smaller buffer gardens will be more likely to be part of, or visible from, the individual's home. Results of some studies suggest that being able to view greenspace from home (e.g. through a window) may confer psychological benefits[45,46], and small buffers (e.g. 250 m) have been used in other UK-based studies that have demonstrated a link with mental health[47,48].

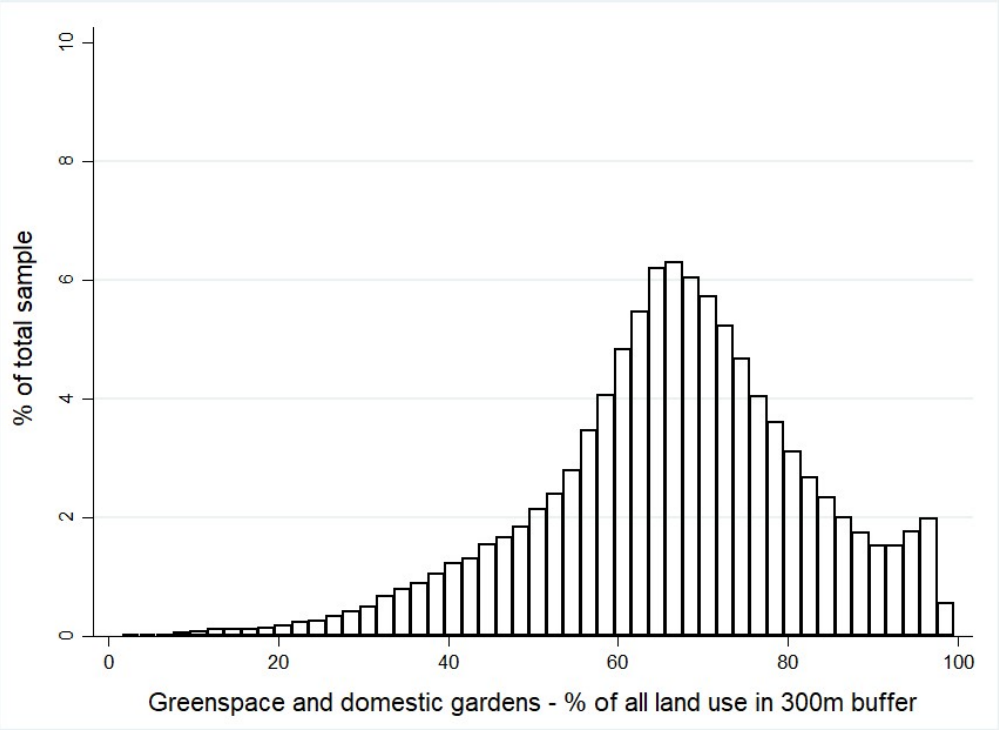


Figure 3.8 Distribution of neighbourhood greenspace (300m buffers)

The mean percentage land use classified as greenspace or garden in a 300m neighbourhood buffer was 66.4% (SD=12.8%) but with a range spanning 1.6% to 99.3% (Figure 3.8). For consistency with the other exposures examined in Chapter 8, I categorised this measure into four ordinal groups, in this case using quartiles (Table 3.7), which were then used as the greenspace exposure variable in Chapter 8.

Table 3.7 Categories of neighbourhood greenspace

Quartile of greenspace (300m buffer)	Mean % green	min	max
Q1 Least green	45.65	1.64	57.88
Q2	62.91	57.88	67.14
Q3	71.49	67.14	76.45
Q4 Most green	85.58	76.45	99.30

As the GLUD measures were available only for those UK Biobank participants resident in England, the analyses using this exposure excluded anyone living in Wales or Scotland. Among residents of England, 1.1% were missing data on greenspace exposure.

In Chapters 6 and 7 I also consider access to 'parks' or 'natural' spaces as a possible effect modifier of associations between the formal PA environment and adiposity, but in those analyses I use different measures more specifically suited to the research questions, and which I detail in Section 3.3.4.2.

3.3.4. Effect modifiers

A central focus of this thesis is to explore possible sources of effect heterogeneity in the associations between the neighbourhood built environments and health. To do this I identified seven potential modifiers of the association between one or more of the primary exposures and one or more of the primary outcomes, each operating at one of three distinct levels: individual, neighbourhood, and macro-environment.

3.3.4.1. Individual-level effect modifiers

At the individual- or intrapersonal- level, a characteristic of an individual may render them more or less susceptible to developing the outcome of interest, in any given environment. Such characteristics would then be considered effect modifiers. I examined the following three potential, individual-level effect modifiers.

Sex/Gender

Effects of neighbourhood characteristics may differ for men and women, with some studies having observed stronger associations among women than among men[49,50]. Differences might be explained by traditional gender roles that result in women spending more time in their local neighbourhood[50]. This is one potential effect modifier of neighbourhood-health associations that has been relatively widely investigated. In Chapter 4, where the primary focus is to first estimate the main cross-sectional associations in the sample as a whole, I also examine possible effect modification by sex. In Chapter 8, where I consider cancer and cardiovascular disease outcomes, I present sex-stratified results alongside combined results, due to different baseline risk among men and women. UK Biobank reports participants' sex (female or male) derived from central NHS registry data, and participants had the option to update this information. No additional data were collected on gender identity.

Table 3.8 UK Biobank participants by sex

Sex	n	%
Female	271,335	54.40
Male	227,411	45.60
Missing	1	0.00
Total	498,747	100.00

Household income

Health inequalities by income and other measures of socioeconomic position are well documented, and there is also some evidence of effect modification by income of neighbourhood effects on BMI[51], obesity[52] and diabetes[53] in the US. It is not clear whether income modifies such associations in the UK, though one recent study suggests a double burden of exposure to fast-food outlets and low income in London[54]. Any neighbourhood characteristics to which residents might be price-sensitive could plausibly be expected to have a directly differential effect by household income (e.g. pay-to-use recreation facilities). Other, less direct mechanisms may also generate effect heterogeneity. Household income is recorded in UK Biobank in five categories of annual household pre-tax income in pounds sterling: less than £18,000, £18,000 - £30,999, £31,000 - £51,999, £52,000 - £100,000, more than £100,000.

Table 3.9 UK Biobank participants by household income

Average household annual income (pre-tax)	n	%
Less than £18,000	97,208	19.49
£18,000 – £30,999	108,180	21.69
£31,000 - £51,999	110,777	22.21
£52,000 - £100,000	86,272	17.30
Greater than £100,000	22,932	4.60
Don't know	21,305	4.27
Refused to answer	2,221	0.45
Missing	49,852	10.00
Total	498,747	100.00

In Chapter 4 I collapsed the highest two income categories so there were four roughly evenly sized groups, and then tested for interactions with availability of formal physical activity facilities and proximity to a fast-food outlet, with respect to adiposity outcome measures. In Chapter 8, a binary indicator of high household income (at least £31,000) was used to investigate income as an effect modifier of associations with CVD and cancer. In

that chapter a binary effect modifier was preferred for the type of analyses being performed (see Chapter 8 for details). Unfortunately, household income was not readily able to be equivalised to account for household size, nor could I take into account housing costs or taxes. As I describe in the later section on confounders, I adjusted all other models for potential confounding by income, using the five-category form of the variable.

Genetic risk

Family and twin studies have shown that BMI is likely to be between 30% and 70% heritable[55–57]. Except in rare cases (e.g. syndromic obesity), it is generally thought that the genetic determinants of body weight are largely polygenic (involving multiple genes), and BMI has been associated with a large number of genetic loci, particularly since the advent of genome-wide association studies (GWAS)[56]. In Chapter 5 I operationalise genetic risk using two polygenic risk scores for obesity, based on the most current evidence from meta-analyses of GWAS[58], and test whether polygenic risk acts as a modifier of the association between neighbourhood characteristics and BMI. I also examine possible effect modification by selected individual single nucleotide polymorphisms (SNPs) strongly associated with BMI and putatively linked to either dietary intake or physical activity behavioural pathways.

Typically, genetic risk scores (GRSs) are derived by summing the number of high-risk alleles present in an individual, usually weighting by known estimates of effect size. A recent GWAS identified 97 SNPs associated with BMI[58]. As shown in equation 1 below, I used these to construct two alternative GRSs by summing the number of BMI-increasing alleles across the set of SNPs, and weighting the allele count at each SNP by its published effect size (β).

$$GRS = (\beta_1 * SNP_1) + (\beta_2 * SNP_2) + \ldots (\beta_n * SNP_n) \qquad \text{(equation 1)}$$

For non-imputed genotypes, the allele count of each SNP takes a value of 0, 1 or 2, while for imputed genotypes, I used the SNP's imputed allelic dosage, which is a value between 0 and 2 representing the sum of the probability of each of the three genotypes at a given SNP.

The first GRS used 91 of the 97 BMI-associated SNPs reported in the recent GWAS, with the exception of six that I excluded on the basis that they were reported elsewhere either as being in linkage disequilibrium with other SNPs in the GRS, or of having pleiotropic effects[59]. Linkage disequilibrium and pleiotropy can introduce bias in the associations

between the genetic risk score and the outcome, and in interaction analyses[60]. To construct the second GRS, I followed another recent study using UK Biobank data and a GRS for obesity[59], in which they also based their SNP selection on the 97 SNPs identified in the recent GWAS[58], but limited their GRS to 69 SNPs identified in the primary meta-analysis of studies of individuals of European descent, excluding the SNPs identified in the secondary meta analyses of studies in regional, sex-stratified or non-European-descent populations. The distribution of the genetic risk scores are shown in Figure 3.9.

To ensure higher values of the GRSs represented increased risk of obesity, it was necessary to recode some of the SNP genotypes so that the non-reference allele in the dataset was the same allele associated in the GWAS with higher BMI. Further details of the GRS derivation are provided in Chapter 5, and full lists of the SNPs included in each GRS can be found in the Supplementary Material for Chapter 5 (Appendix Two).

Separately I also examined six individual SNPs as possible effect modifiers. These were identified from the literature as having well-established links to obesity and food intake (markers of the *FTO*, *MC4R* and *TMEM18* genes)[57,58], or had been linked specifically to physical activity (markers of *CADM2*, *GNPDA2 NRXN3*)[61,62]. Full details are provided in Chapter 5. If any of these individual SNPs do interact with the food or physical activity environment to influence BMI, it would be expected that those linked to diet would interact only with the food environment, while those linked to physical activity would interact only with the physical activity environment.

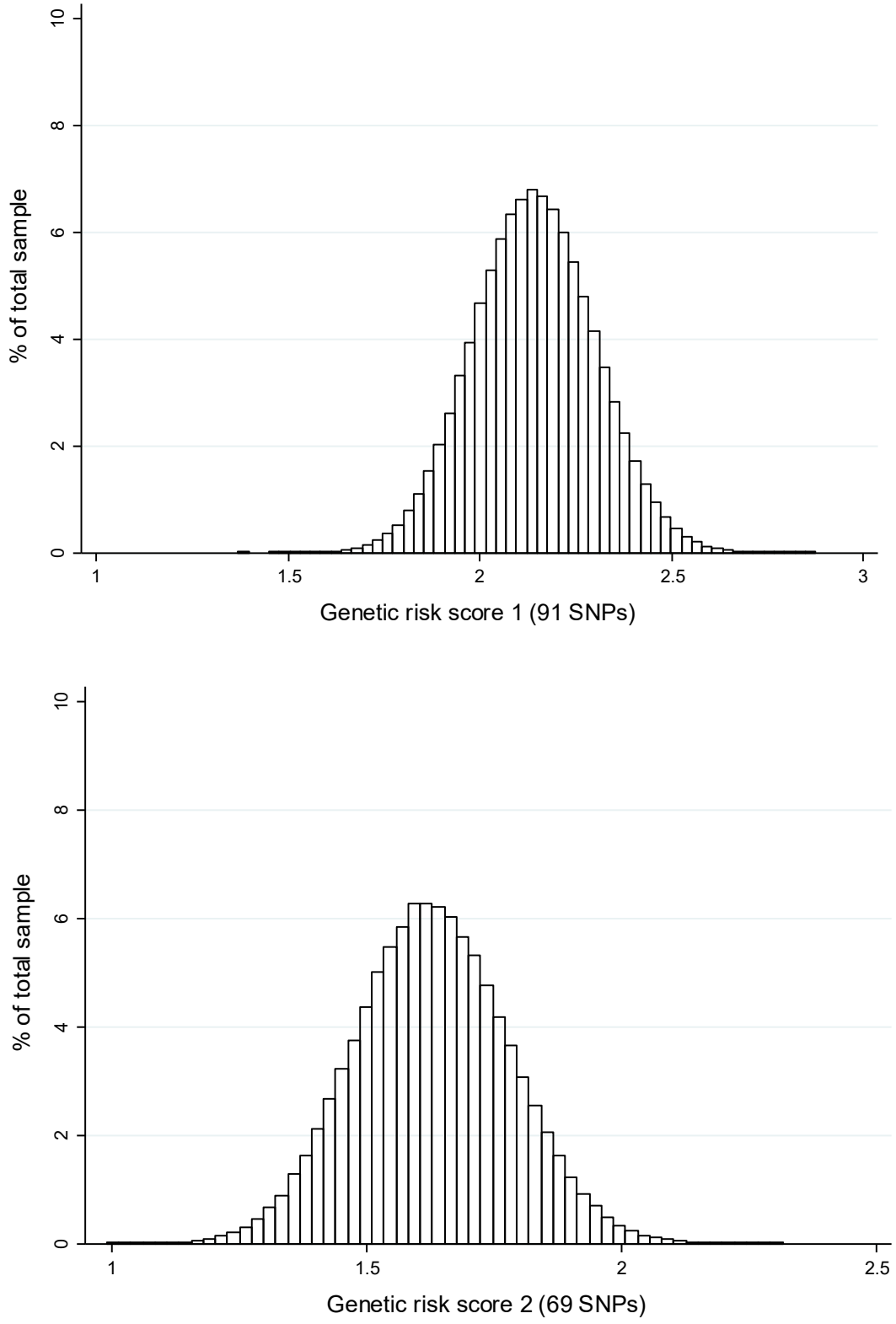


Figure 3.9 Distribution of genetic risk scores for BMI in UK Biobank cohort

3.3.4.2.Neighbourhood-level effect modifiers

At the level of neighbourhood, loosely defined, physical or social characteristics of a neighbourhood may modify the effect of another neighbourhood characteristic on an outcome of interest. In Chapter 6 I examine two characteristics of the built environment as potential, neighbourhood-level effect modifiers of the relationship between neighbourhood availability of formal PA facilities and adiposity.

Availability of informal physical activity resources (parks and other public open/green spaces)

Studies of neighbourhood physical activity environments and obesity tend not to routinely consider how co-occurring characteristics of the built environments might act in synergistic or antagonistic ways to influence physical activity and health. For example, formal physical activity facilities such as gyms and swimming pools may play an important role for people living in areas with few informal resources that promote physical activity (such as parks and other public spaces). To examine this hypothesis, in Chapter 6 I consider whether the number of parks or other public open/green spaces in a one kilometre street-network buffer around each participant's home acts as an effect modifier of the association between the availability of formal physical activity facilities and adiposity.

This variable is derived from the UKBUMP data on land-use densities, in the same way I derived the formal physical activity facilities measure, but in this case counting the number of sites categorised as:

- Park
- Public Park/Garden
- Public Open Space/Nature Reserve
- Playground
- Open Space/Heath/Moorland.

Because the distribution was highly positively skewed, and to reflect the measure of formal physical activity facilities, I categorised this as zero, one, or at least two parks or other informal physical activity resources within 1 km.

Table 3.10 UK Biobank participants by category of neighbourhood availability of parks and other public open/green spaces

Number of parks and other public open/green spaces within 1km	n	%
0	215,401	43.19
1	91,213	18.29
2 or more	174,453	34.98
missing	17,680	3.54
Total	498,747	100.00

Proximity to fast-food outlets

Just as the absence of informal physical activity resources might strengthen the influence of formal physical activity resources on obesity risk, so might an unhealthy food environment have the opposite, antagonistic effect, overriding the potentially health-promoting influence of the neighbourhood physical activity environment on energy balance and resulting adiposity. Chapter 6 therefore also includes examination of the fast-food environment as a possible effect modifier of the formal physical activity environment. I used the measure of proximity to a fast-food outlet that I had previously defined for analysis as an exposure in other chapters, but collapsed to three categories rather than four, for internal consistency within that chapter (so, <500 m, 500-1499 m, or at least 1500 m from home to the nearest takeaway/fast-food outlet).

Table 3.11 UK Biobank participants by category of proximity to nearest fast-food outlet

Distance to nearest fast-food/takeaway outlet	n	%
Closer than 500m	88,789	17.80
500m – 1499m	208,308	41.77
At least 1500m	183,923	36.88
missing	17,727	3.55
Total	498,747	100.00

Area-level deprivation

Neighbourhood deprivation has been shown in many studies to be associated with poorer health outcomes, probably operating through various complex mechanisms[63–65]. Accordingly, studies of built environments and health usually adjust for deprivation, or concentrate their focus on deprived neighbourhoods. While it has been shown not to be universally the case[66], there are many examples of more deprived areas having lower quality neighbourhood resources – a process referred to as 'deprivation amplification'[67]. One dimension of this phenomenon is the possibility that while broad measures of the neighbourhood built environment might not reveal greater exposure to unhealthy resources, the (often unmeasured) quality of some available resources may be worse, for example poorer quality public green spaces, and perceptions of the built environment might be affected by issues of safety or crime[68–70]. On the other hand, if deprived areas do have fewer healthy resources but only in one domain of the built environment (e.g. formal PA facilities) the effect of other domains (e.g. informal resources for physical activity, such as greenspace) might be stronger there than in less deprived areas that are well resourced across all domains. It is therefore difficult to predict which direction any effect modification by area deprivation would operate in, and it is also possible that an absence of statistical evidence for effect modification by area deprivation could arise from multiple processes operating in opposing directions..

Attempting to overcome some of these challenges for interpretation, in Chapter 8 I investigate whether the associations of three distinct characteristics of the neighbourhood built environment with hospitalisations due to CVD or cancer are modified by area deprivation. This is measured using the Townsend score of each participant's census output area. Census output areas are statistical units of varying sizes, the majority (80%) comprising between 110 and 139 households and therefore representing an area fairly local to a person's home. The Townsend deprivation index incorporates four variables: unemployment (% aged 16 and over who are economically inactive); non-car ownership (% of all households); non-home ownership (% of all households); and household overcrowding. These variables are measured from the 2001 census for each census output area and combined (via a series of calculations involving log transformations and standardisations) to give a 'Townsend score' for that output area. A greater Townsend index score implies a greater degree of deprivation. UK Biobank participants were assigned a score corresponding to the output area in which their postcode is located. Using publicly available quintile boundaries for the 2001 Townsend index[71], I constructed a binary indicator of deprivation (areas in the three least deprived quintiles vs those in the two

most deprived quintiles) to investigate area deprivation as an effect modifier. This is done alongside the investigation of household income as an effect modifier for the same associations (described earlier), and each adjusted for confounding by the other, to attempt to isolate modifying effects of each. All other models in the thesis were adjusted for confounding by area deprivation, as described in Section 3.3.6.

Table 3.12 UK Biobank participants by area deprivation

Quintile of Townsend 2001 area deprivation score	n	%
Least deprived areas	211,466	42.40
2	67,715	13.58
3	65,313	13.10
4	77,024	15.44
Most deprived areas	76,607	15.36
missing	622	0.12
Total	498,747	100.00

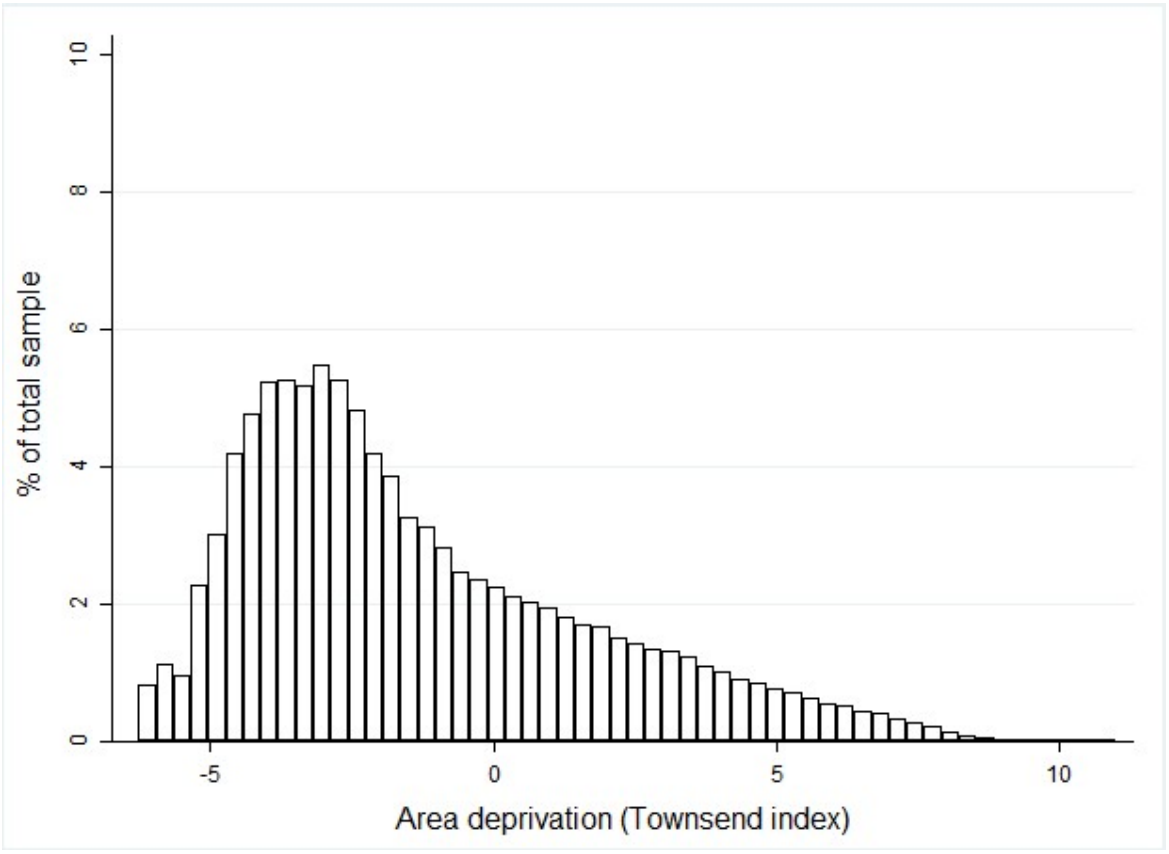


Figure 3.10 Distribution of area deprivation (Townsend 2001 score) in UK Biobank cohort

3.3.4.3.Macro-environmental effect modifiers

Studies of neighbourhoods and health across a range of settings have yielded inconsistent findings, indicating possible geographical heterogeneity in the health effects of neighbourhood characteristics. This in turn suggests that macro-environmental attributes of larger geographical areas within which neighbourhoods are located might play a modifying role. Understanding any such drivers of geographical effect heterogeneity may help clarify observational findings and inform local policy interventions. I examined the following two potential effect modifiers at the level of Local Authority District:

Percentage of land cover classified as 'natural'

People living in cities and towns that are surrounded by accessible natural landcover (woodland, moors, beaches, etc.) have enhanced informal opportunities for outdoor physical activity even if those natural spaces are not within one's immediate neighbourhood. Increased exposure to such environments in the wider area may also contribute to a local culture of outdoor recreation. In such places, there may be less reliance on, or normalisation of, using formal physical activity facilities such as gyms and leisure centres close to home. In such places, we might therefore see a reduction in the magnitude of association between the neighbourhood availability of formal physical activity facilities and adiposity.

The percentage of land cover classified as 'natural' in each Local Authority in England is compiled in the Land Cover Atlas of the UK[72], having been derived from Corine Land Cover data from 2012 (as described in Chapter 2). The 'natural' land cover classification includes all land cover that is neither 'artificial' (urban, industrial, commercial, transport, mining etc.) nor 'agricultural'. The 'natural' classification covers land cover types such as forests, grasslands, moorland, beaches, wetlands, and water bodies. It does not include farmland such as pastures, which is classified as 'agricultural' or urban green areas such as parks and sport and leisure facilities (e.g. playing fields), which are classified in Corine as 'artificial' and in the Land Cover Atlas of the UK as their own category of 'urban green'. Corine Land Cover data are based on remotely-sensed satellite imagery spanning the whole of Europe, which classified land use based on the identification of large (≥25 hectares), relatively homogeneous areas. More heterogeneous areas are classified depending on the dominant land use; for example, 'discontinuous urban fabric' areas will contain a mix of buildings and artificial surfaces, but also smaller vegetated areas, while large urban parks or urban woodlands will be classed as 'green urban areas'. Thus, the 'natural' classification refers

only to large areas of natural land cover, while smaller green spaces are classified elsewhere.

Figure 3.11 shows the distribution of natural landcover percentage across the 158 Local Authorities in England represented in UK Biobank. Natural landcover percentage ranged from 0.0% to 55.6%, with a median of 5.1% (IQR: 1.2-12.2%). Due to its positive skew, natural landcover percentage was square-root transformed prior to analysis. As people living in rural areas may have a different relationship to the natural environment[73–75], I restricted the analysis to the 86% of the UK Biobank cohort living in areas that are classified by the Office of National Statistics as urban (specifically, where a person's home postcode is located within a city or a town that has a population of at least 10,000 people).

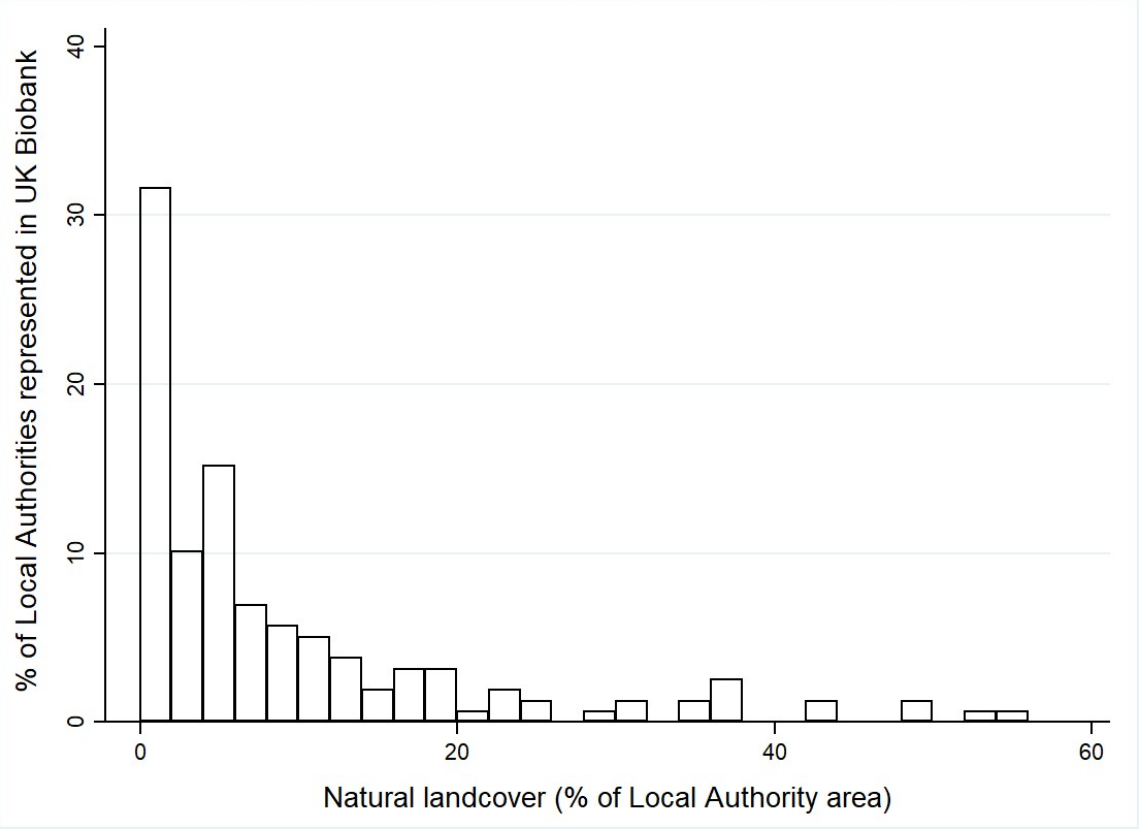


Figure 3.11 Distribution of 'natural' landcover percentage in Local Authorities in England represented in UK Biobank

Local descriptive norms for obesity

Spatial variation in the prevalence of particular traits (e.g. obesity) or behaviours (e.g. PA) creates what are known as local descriptive social norms[76]. I hypothesise here that in areas where adult obesity prevalence is high, social pressure to be a healthy weight may be weaker due to obesity becoming 'normalised'[77], and this may attenuate any health-promoting influence of the neighbourhood built environment on adiposity. To test this, I use Local Authority-specific estimates of obesity prevalence among adults (aged 16 or over) for the period 2003–05[78]. This period immediately precedes the recruitment for UK Biobank, so approximates the social norms that prevailed at the time of the baseline assessment (without relying on the same data source).

Obesity prevalence estimates were only available for LADs in England, so the analysis using this variable (Chapter 7) was restricted to UK Biobank participants residing in England. This has the advantage of also reducing the risk of confounding due to contextual differences that might arise from historical or current differences between the devolved nations of the UK.

Figure 3.12 shows the distribution of estimated obesity prevalence across the 158 Local Authorities in England represented in UK Biobank. Mean obesity prevalence was 23.0% (SD=3.7%). An estimate of obesity prevalence was unavailable for seven Local Authorities so for these I used regional estimates of obesity from the Health Survey for England for the same period, imputing the value for the region in which each Local Authority was located. I treated obesity prevalence as a continuous variable.

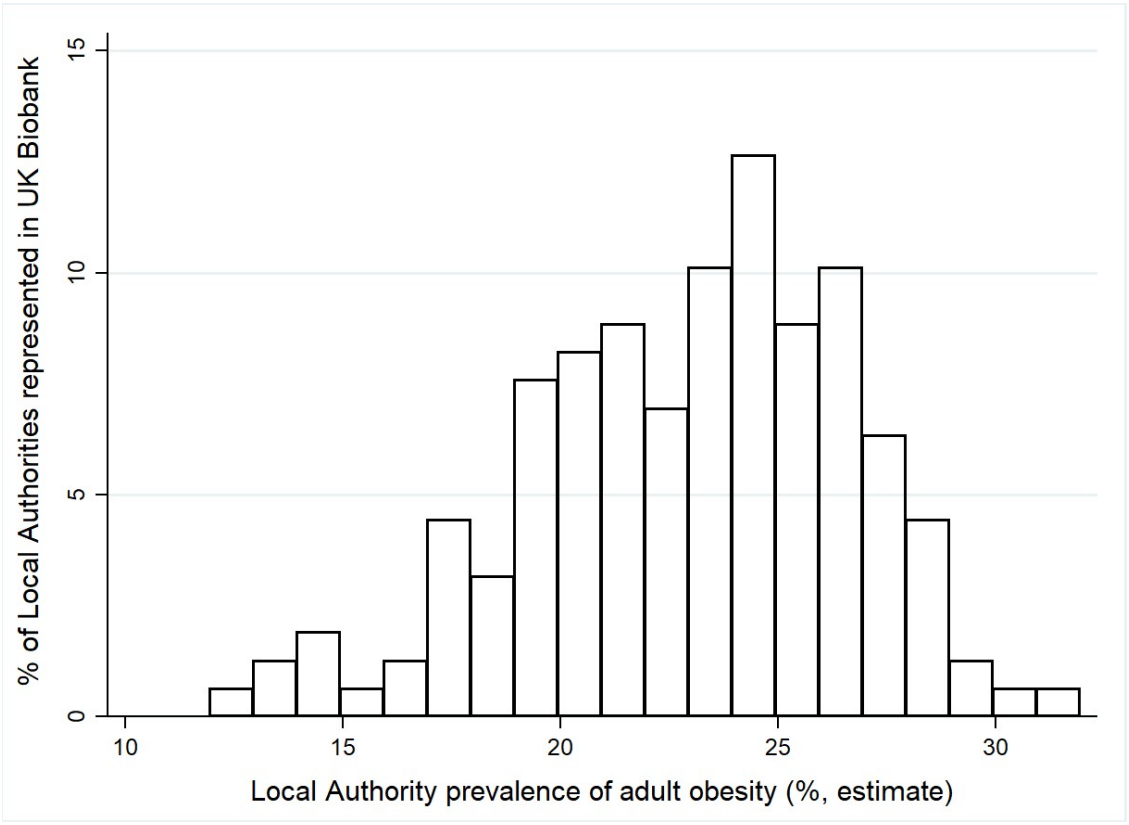


Figure 3.12 Distribution of estimated adult obesity prevalence in Local Authorities in England represented in UK Biobank

3.3.5. Covariate selection

3.3.5.1. Directed Acyclic Graphs to identify potential confounders

Based on the existing literature and careful consideration of the conceptual model underpinning the relationships under investigation in this thesis, I identified potential confounders of each association of interest. This information was then summarised in directed acyclic graphs (DAGs) to assess which variables should be adjusted for in each analysis. This also enabled me to identify any situations where adjustment for a potential confounder might induce bias through another pathway. I used the web-based DAGitty interface to draw the DAGs and help to identify the minimally sufficient adjustment sets[79]. Supplementary Figures 1 and 2 of the published paper in Chapter 4 show DAGs for the relationships between the physical activity environment and adiposity, and the fast-food environment and adiposity (Appendix One).

As illustrated in these DAGs, potential confounders included individual demographic and socioeconomic variables and several local area characteristics. Each paper in the thesis contains a brief justification for the adjustment or not for specific variables. In general,

alongside socio-demographic characteristics of the individual that could confound estimates, models are adjusted for area-level deprivation as a strategy to minimise confounding by other unmeasured neighbourhood characteristics[80]. Physical activity environment and fast-food environment models are also mutually adjusted for one other. The operationalisation of all covariates used in the thesis is described in a subsequent section of this chapter.

3.3.5.2. Checks for multicollinearity

Having used DAGs to articulate the likely confounding relationships and identify the minimally sufficient adjustment sets, I then used change-in-mean squared error (ΔMSE) procedures to formally check that adjusting for potential confounders using the observed variables in UK Biobank did not introduce multicollinearity into the regression models[81]. Briefly, this involves comparing a fully adjusted model to a series of models in which each covariate is dropped, to check that the inclusion of any one covariate doesn't inflate the mean squared error of the main effect estimate – if it does, this indicates that the reduction in bias gained by including the variable is outweighed by the inflation of the standard error of the effect estimate. For simplicity, I did this using OLS (rather than mixed effects) regression models and assumed the conclusions drawn from these checks would apply across the related but more complex models used in this thesis. The change-in-MSE checks indicated multicollinearity was not a problem and it was appropriate to proceed with the adjustment sets identified using the DAGs.

3.3.5.3. Controlling for confounding in analyses of effect modification

As mentioned earlier, effect modifiers need not themselves be direct causes of the outcome. If they do cause the outcome, we might be interested in either statistical interaction (effect modification) or biological interaction. The latter exists when the effect of the exposure on the outcome is biologically dependent on the presence or absence of a third variable, rather than that effect simply varying across levels of the third variable, without being a necessary condition for causing the outcome. Effect estimation in the context of biological interaction requires that confounders of both exposure-outcome relationships be adjusted for, while in the case of effect modification (statistical interaction), one need only adjust for confounders of the primary exposure-outcome association to estimate an unbiased effect[82]. As I am only concerned with effect modification in this thesis, I take the latter approach to covariate adjustment in my regression models.

3.3.6. Operationalisation of potential confounders

3.3.6.1. Socio-demographic covariates

The following variables measuring demographic and socioeconomic characteristics of participants are included as covariates in the primary analyses, to control for potential confounding: **age**, **sex**, **highest level of education attained**, **annual household income, employment status, ethnicity.** These might act as confounders because they tend to be associated with the outcomes (e.g. average BMI is higher among older people, Black/Black British ethnic groups, and people with lower educational attainment), while also influencing the type of neighbourhood a person resides in, via affordability, lifestyle preferences, and residential segregation along ethnic or religious lines.

Age was treated as a continuous variable centred around its mean. Nine cases were excluded as their age at baseline assessment was recorded as being <40 or >70. These exclusions were made because 40-69 years was the eligible age when identifying the sample to be targetted for recruitment. Since assessment appointments were scheduled for several weeks after the initial recruitment letter, some individuals may have turned 70 by the time of assessment. The reasons for the small number of participants outside this range is not explained in the UK Biobank documentation.

Sex was included as a binary variable (male or female).

Highest level of **education** was classified as College/University degree, A/AS levels or equivalent, O levels/GCSEs or equivalent, CSEs or equivalent, NVQ/HND/HNC or equivalent, Other professional qualifications, or None of the above, and included as a categorical variable.

Income was included in models as the following categories of annual household pre-tax income in pounds sterling: Less than £18,000, £18,000 - £30,999, £31,000 - £51,999, £52,000 - £100,000, Greater than £100,000. In the models in Chapter 8 where income was treated as a confounder (rather than a potential effect modifier as it was in some of that chapter's models) the top two categories were collapsed because of concerns about data sparsity in the proportional hazards models.

Employment status: multiple responses were allowed, to capture a variety of roles. The following hierarchy was imposed when multiple employment statuses were identified: paid work, retirement, unable to work, unemployed, other. The 'other' category includes home/caring duties, volunteer work, student, and 'none of the above', due to small numbers of people identifying these as their sole employment status (when other

categories were also selected, those were given precedence). Although this approach appears to devalue unpaid work such as home and caring duties, it has been taken here so that the variable provides the best reflection of status in the paid labour force, and with a particular interest in distinguishing between people more likely to be exposed to workplace and commuting environmental influences, and those who may have proportionately greater exposure to their home/neighbourhood environment.

Detailed information on self-reported **ethnicity** was collected at assessment, and for the purposes of its inclusion as a covariate in these analyses it has been reclassified into eight categories: White; South Asian/South Asian British; Black/Black British; Chinese/other Asian; Mixed: White/Black; Mixed: White/Asian; Mixed – no detail; Other. This differs from the standard ONS-recommended harmonized ethnicity coding because the ONS scheme groups together ethnic groups known to have different BMI-health relationships (specifically, South Asian is combined with Chinese and other East Asian groups, and Mixed covers any combination of ethnicities). I have instead split the 'mixed' classifications up by the non-white component since that is specified in many cases in this sample, and generated an 8-category variable that more appropriately captures the nuance otherwise lost in coarser groupings of 'mixed' and 'Asian'. Exploratory analyses suggested the primary analyses were not sensitive to the classification scheme used.

3.3.6.2.Area-level covariates

Three measures relating the area in which participants lived were also included as covariates in primary analyses. These were: **area deprivation, urban/non-urban status** (except in analyses restricted to urban residents), and **neighbourhood residential density**. These area-level measures are considered potential confounders because various neighbourhood characteristics tend to cluster together (e.g. more fast-food outlets in deprived areas, more physical activity resources in densely populated, urban areas), and they capture dimensions of neighbourhoods that are also associated with the health outcomes of interest in this project. Models in Chapter 7 are also adjusted for the **gross disposable household income (GDHI) per capita** of the Local Authority in which they live, to control for possible confounding effects of the wider socioeconomic context.

Area deprivation was measured by the Townsend score (see earlier description in Section 3.3.4.2 for details). When modelled as a covariate it was treated as a continuous variable.

Urban/non-urban status was included as a binary indicator based on postcode, whereby all postcodes located within an area with population of at least 10,000 people were classified as urban and all others as non-urban. The primary analyses reported in Chapters

6 and 7 were restricted to the urban subsample and therefore those models were not adjusted for this.

Residential density: Using the same land-use density dataset in UKBUMP as for the physical activity environment measures, the density of residential land use around each person's home address was measured as a count of all features within a 1000 m street network buffer that are classified as residential (including various types of dwelling as well as garages/parking spaces) in the UK Ordnance Survey AddressBase database. As this measure is highly negatively skewed, I $\log_{10}$ transformed it before inclusion in analyses.

Gross disposable household income (GDHI) per capita (for 2006), was used in Chapter 7 to control for possible confounding effects of the wider socioeconomic context[83]. GDHI is the amount of money that all of the individuals in the Local Authority District have available for spending or saving after any income and benefits have been received and direct and indirect taxes have been paid. GDHI is a concept that is seen to reflect the collective 'material welfare' of individuals in a region[83].

Additionally, in single-level models (Chapter 8), **assessment centre** was included as a covariate, while in multilevel models (Chapters 4-7) this was included as the level-2 identifier rather than a covariate. This is described further in Section 3.4 on statistical methods.

3.3.6.3.Other covariates

In Chapter 8, proportional hazards models for CVD- and cancer-related hospitalisations are additionally adjusted for **smoking status** (current/previous/never), **alcohol intake frequency** (less than/at least 3 times per week) and **number of years living at current (baseline) address** (continuous). I dichotomised alcohol intake frequency from the original 5-point scale on which the data were collected, after finding that results were insensitive to how the variable was classified.

3.3.6.4. Mutual adjustment for other neighbourhood exposures

Additionally, in Chapters 4, 5 and 7, where there are two primary environmental exposures (fast-food proximity and availability of physical activity facilities) each is also treated as a potential confounder of the association between the other environmental exposure and adiposity, since the two are correlated. That is, in models of the PA environment and adiposity, distance to nearest fast-food outlet was included as a covariate, and in models of the fast-food environment, availability of formal PA facilities was included as a

covariate. This was also undertaken for models of the physical activity environment in Chapter 6.

3.3.6.5.Additional covariates included in sensitivity analyses

Chapter 4 includes sensitivity analyses to test whether the association between the physical activity environment and adiposity is confounded by dietary intake, and whether the association between the fast-food environment and adiposity is confounded by physical activity. This was done by including the following additional covariates in regression models:

Total energy intake (KJ): This was collected via 24-hour recall dietary assessment from a subset of respondents. Dietary assessment was added to the baseline assessment protocol late in the recruitment phase, and was therefore collected for only ~70K participants at the baseline assessment visit. After the close of recruitment, 4 additional questionnaire rounds were conducted online, with invitations emailed to participants at 3-4 monthly intervals. This is therefore after the baseline visit when anthropometry measurements were collected. Involvement by participants was voluntary, and participants could respond at multiple time points, so there are multiple recordings for some people. Ultimately, responses were collected from a total of 210,140 unique individuals between April 2009 and June 2012. The variable I constructed for analysis uses the earliest available measurement for anyone with any measurement recorded.

Total physical activity: Data on the type and duration of physical activity were collected from UK Biobank participants via the touchscreen questionnaire at the baseline assessment, using the short form of the International Physical Activity Questionnaire (IPAQ). Participants were asked to report on the number of days in a typical week they engaged in walking, moderate physical activity or vigorous physical activity, and for each type of physical activity engaged in at least once in a typical week, how many minutes they typically engaged in these activities per day. Following the IPAQ guidance, this information was then combined and weighted by the energy requirements (in metabolic equivalents (METs)) of each activity, to derive a measure of total energy expended through self-reported physical activity, expressed in terms of MET minutes per week. As this measure is highly positively skewed, it was recoded to a categorical variable indicating low, moderate or high levels of physical activity, following IPAQ guidance.

Chapter 8 includes sensitivity analyses to test whether the models of the associations between characteristics of the neighbourhood environment and hospital admissions due to CVD and cancer are confounded by **baseline BMI, hypertension** or **medications for**

hypertension or cholesterol. BMI was included as a continuous variable and hypertension and medications as binary indicators.

3.3.6.6. Adjustment for latent genetic structure

It has recently been shown that there exists apparent latent genetic structure in the UK Biobank sample that, contrary to expectation, is not accounted for using routine model adjustment for assessment centre and principal components derived from the genotype data[84]. To explore the possibility that my results in Chapter 5 might be biased by this latent genetic structure, I performed sensitivity analyses in which models were adjusted for all available genetic ancestry principal components provided by UK Biobank (rather than the standard approach of adjusting for the first 10), and for geographical coordinates of birth location.

3.4. Statistical methods

3.4.1. Treatment of missing data

The primary environmental exposure variables were missing for up to 3.6% of the relevant sample, as reported earlier.

Of the primary covariates (excluding diet and physical activity), income has the greatest degree of missing data (15%, of which 10% refused to answer, while 4.3% didn't know and 0.5% were missing for another reason). All other individual-level covariates are missing at a frequency of 1% or below.

Thus, the main concern is whether income was missing in a way that would systematically bias the results. Principally, whether the exposure-outcome associations are different among participants who did not report income data compared with the rest of the sample.

In preliminary work for the paper in Chapter 4, I compared excluded cases to the complete case sample. Comparisons across covariates showed this group to be from less deprived areas, more likely to be retired or otherwise not in the labour market (e.g. home duties), less highly educated, and more likely to be of South Asian or Black ethnicity. I then compared the coefficients from models not adjusted for income with those from fully adjusted models, and observed that excluding income from the models increased the coefficients by around 10% at most (both in models including cases missing on income, and those without them). I therefore determined that the effect of income missingness was unlikely to spuriously inflate estimates. On this basis, along with the fact that on theoretical grounds it would be hard to justify not including income as a covariate, and

the fact that examination of effect modification by income was one of my research questions, I decided to retain the adjustment for income despite the missing data issue, and analyse the data for complete cases, including on income, only.

Implications of any substantial missing data on variables only used in sensitivity analyses are discussed in the relevant chapters (in particular Chapter 4).

3.4.2. Descriptive statistics

Beyond the distributions and other summary statistics reported in the preceding description of the key variables, descriptive statistics for all key variables are reported within each empirical chapter as they relate to that analysis.

Continuous variables are summarised by their mean and standard deviation if they are approximately normally distributed, and by the median and interquartile range otherwise. Categorical variables are summarised by the number and percentage in each category.

3.4.3. Multilevel regression modelling

The general analytical approach I have taken in this thesis is to use multivariable regression models to estimate independent associations between each characteristic of the neighbourhood environment and each outcome, and then investigate modification of those associations by including interaction terms in the models.

Due to the clustered sampling design of UK Biobank, the data has a hierarchical structure, whereby individuals are nested within assessment areas. To account for this structure, I use multilevel (mixed effects) models in Chapters 4, 5 and 6, specifying assessment centre as the level-2 cluster identifier, to analyse cross-sectional relationships. In Chapter 7, I am explicitly interested in the nesting of individuals within Local Authority Districts, so in that case I use Local Authority instead of assessment areas to identify clusters. In Chapter 8, when analysing time-to-event data, I use single-level proportional hazards models with a dummy variable for each assessment area.

In the first research paper (Chapter 4), I estimate multilevel, multivariable linear regression models with random intercepts, and random coefficients for the environmental exposures, and I specified an unstructured variance/covariance matrix. Specifying a random intercept allows the mean value of the outcome to vary across assessment centres, and specifying a random coefficient for the main exposure variable allows the exposure-outcome association to also vary by assessment centre, thus acknowledging the possibility that the association of interest may not be uniform across all study sites. Assuming an

unstructured variance/covariance matrix allows the random intercepts and coefficients to covary[85]. Covariate adjustment sets were determined for each model as earlier described (p.93). Results were expressed as the mean difference in adiposity associated with a unit/categorical change in the exposure, and 95% confidence intervals (95% CI). This same model is used as a basis for the analyses in Chapters 5-7.

While I am not specifically concerned here with estimating or interpreting variance at the level of assessment area, this modelling approach aligns theoretically with the conceptual framework underpinning this thesis: that is, that these neighbourhood-health associations may not be uniform across geographical space or the population. I formally tested whether multilevel models were a better fit to the data than equivalent OLS models, by comparing the OLS to the random intercept model using the likelihood ratio test. Upon determining that the multilevel random intercept model was preferable, I then compared this to the model with random coefficients for the exposure, and these too improved model fit.

In Chapter 7, where I investigate geographical heterogeneity across Local Authority Districts of England, the first stage of the analysis entails additionally estimating single-level, OLS models for each Local Authority. These are linear models with the same covariate set as the multilevel models.

In secondary analyses in Chapter 4 to examine effect modification by sex and household income, and then for the effect modification analyses that are the focus of Chapters 5-7, I add various interaction terms to this basic multilevel model specification. Each interaction term represents the cross-product of the primary neighbourhood exposure and the potential effect modifier being examined.

In Chapter 7, in that paper's second stage of analysis, I test for cross-level interactions between the neighbourhood characteristic and potential modifiers operating at the Local Authority level. To obtain estimates with a meaningful interpretation I centre the level-1 exposure variable (neighbourhood characteristic) around the cluster (Local Authority) mean of that variable. According to Enders and Tofighi[86] this is appropriate for research questions where a level-1 exposure, and modification of its effect by a level-2 variable, are of substantive interest, as they are in Chapter 7.

In Chapter 8 I examine associations between neighbourhood exposure and incident hospital admission due to CVD or cancer after baseline, using multivariable Cox proportional hazard models. Results are expressed as adjusted hazard ratios (HRs) and 95% confidence intervals (95% CI). I test the proportional hazards assumption by visual

inspection of adjusted log-log plots to detect non-parallel curves (using the `stphplot` command in Stata).

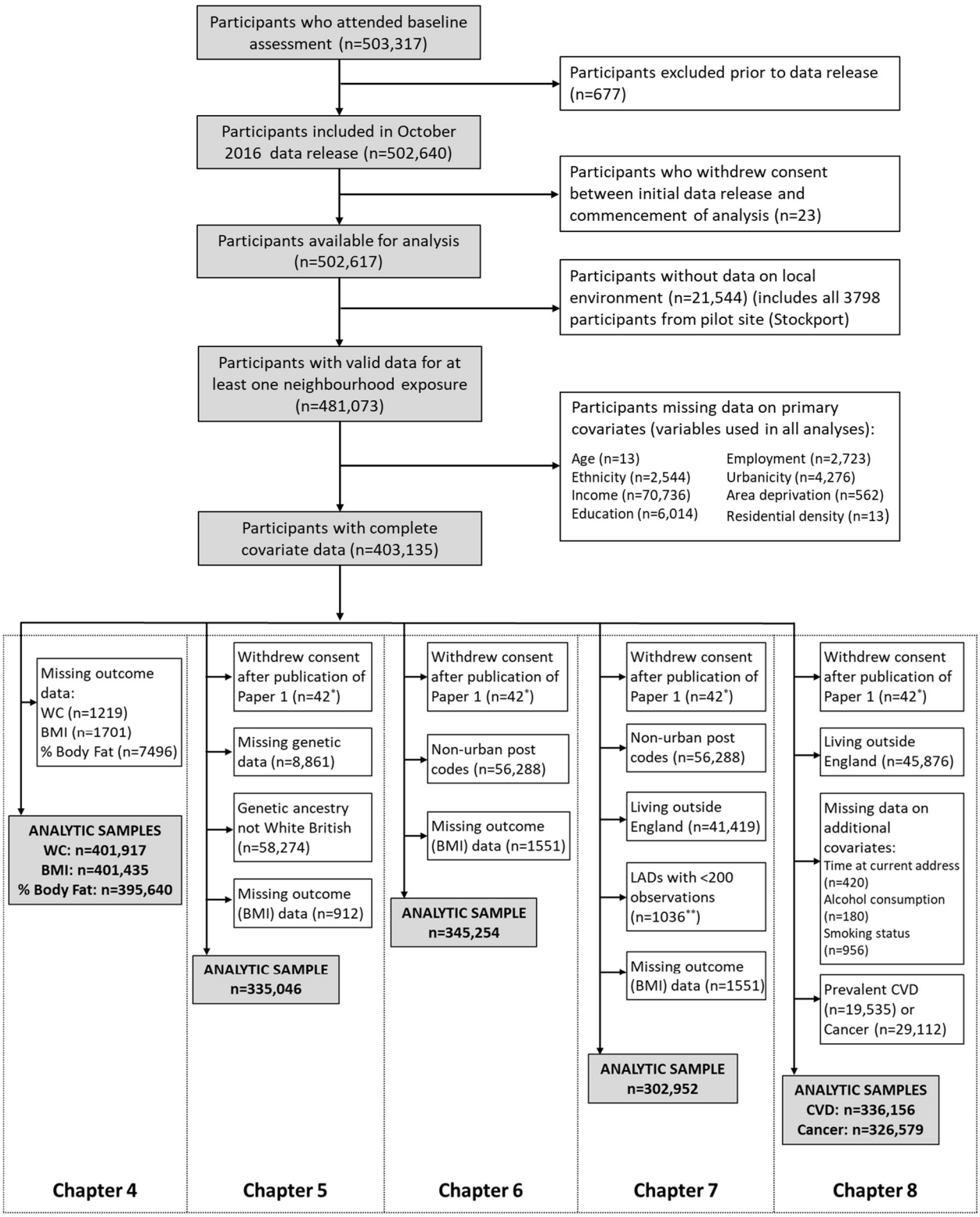
Across the thesis, I focus on estimating the health benefits associated with exposure to theoretically health-promoting neighbourhood characteristics, rather than health deficits associated with exposure to less healthy environments. In practical terms, this means taking the theoretically least health-promoting level of exposure as a referent and estimating relative to that group the differences in adiposity (Chapters 4-7), and relative hazard of admission to hospital (Chapter 8), for those exposed to theoretically healthier environments.

Specific details of each model are provided in the methods sections of the relevant chapters herein. All models for any given outcome were restricted to participants with complete data for all covariates. Final analytical sample sizes therefore varied accordingly. The flow diagram in Figure 3.13 shows the various exclusions and analytical sample sizes for the primary analyses in Chapters 4 to 8.

3.4.4. Assessment of effect modification

When investigating effect modification, it is important to note that it can be assessed on either the additive or the multiplicative scale[2]. The interpretation of effect modification differs, often importantly, depending on which scale is used.

When dealing with continuous outcomes, interaction terms in linear models automatically provide a direct way to assess effect modification on the additive scale. On the other hand, when dealing with binary and time-to-event outcomes, interaction terms in logistic and other non-linear models directly assess effect modification on the multiplicative scale. When an exposure is continuous or binary (as opposed to categorical) the *p*-value for the interaction term in a linear regression model can be interpreted as strength of evidence against the null hypothesis of no interaction on the *additive* scale, while in non-linear models the *p*-value refers to an interaction on the multiplicative scale. If the exposure is categorical, likelihood ratio tests can be used to compare models with and without the interaction term, with these being interpreted on the same scale as the *p*-value for a continuous exposure would be. In all cases, stratified models or marginal predictions then allow assessment of how the effect estimates actually differ across strata of the effect modifier.



* A further 39 also withdrew consent but were already excluded from analysis due to missing data.

** Includes 26 observations with apparently incorrect coordinates (from a total of 91, the rest of which were excluded earlier due to missing data)

Figure 3.13 Flow diagram of exclusions from full UK Biobank samples for analysis

For both continuous and non-continuous outcomes, more complicated techniques can be employed to switch to the other scale for the assessment of effect modification[82]. For continuous outcomes, log-linear models can be used to assess effect modification on the multiplicative scale, and for binary or time-to-event outcomes, measures such as the relative excess risk due to interaction (RERI) can be calculated[87,88]. Presentation of effect modification results from observational studies using RERI measures is recommended by the STROBE guidelines[89].

The additive scale is the more informative with respect to the potential public health consequences of the exposure for different strata of the effect modifier, because unlike the multiplicative scale, it estimates the absolute increase in risk taking into account baseline risk in each strata of the effect modifier[82]. Put another way, the multiplicative scale might indicate larger effects of an exposure in one group than another, but if the risk of the outcome is much lower in that group to begin with, the overall public health consequences of intervening on the exposure in that group will be minimal. On the other hand, the multiplicative scale might show an equally strong effect of an exposure on all groups (i.e. no effect modification), but if one group has a much higher baseline risk of the outcome, intervening on the exposure in that group would benefit a greater number of people. Assessing effect modification only on the multiplicative scale would preclude us from reaching that important conclusion.

Therefore, on the understanding that the additive scale is the preferred scale for assessing effect modification for this thesis, the type of outcome being modelled in each Chapter has dictated how I have assessed effect modification. In Chapters 4-7, where the outcomes are continuous measures of adiposity, I use linear models and am therefore working on the additive scale. Amongst these analyses, when the exposure and effect modifiers are categorical (Chapters 4 and 6), I use likelihood ratio tests to compare linear models with and without inclusion of interaction terms. When both are continuous variables (Chapters 5 and 7), I simply assess the *p*-value for the interaction term directly from the model.

In Chapter 8 where I deal with time-to-event outcomes, I assess effect modification on the additive scale by calculating the RERI. Details of how this is calculated and interpreted are provided in the methods section of Chapter 8. Essentially, the model is estimated with a single reference category for all combinations of levels of the exposure and potential modifier, and the RERI is calculated as per equation 2:

$$RERI = HR_{11} - HR_{10} - HR_{01} + 1 \qquad \text{(equation 2)}$$

where HR_{11} represents the hazard ratio (relative to the reference category) for people 'exposed' to both the primary exposure and the effect modifier, HR_{10} represents the hazard ratio for people exposed to the primary exposure but not the effect modifier, and HR_{01} represents the hazard ratio for people exposed to the effect modifier but not the primary exposure. Evidence of effect modification is indicated by a departure of the RERI from zero (whether above or below zero depends on how the exposure and outcome are defined).

3.4.5. Sensitivity analyses

To test the robustness of my findings to potential sources of bias and various analytical choices that may have influenced the internal validity of the study, in each chapter I include several sensitivity analyses. These are summarised as:

3.4.5.1. Negative control analysis to examine risk of residual confounding

To assess the possibility that the main associations of interest in the thesis are residually confounded, I designed a negative control analysis in Chapter 4 in which I used height as a negative outcome control.

The use of negative controls has been advocated as a way to check for residual confounding[90]. Negative controls can be either exposure or outcome controls, and should be selected such that they have no plausible causal relationship with the outcome (in the case of negative exposure controls) or the exposure (in the case of negative outcome controls), but the control and the exposure or outcome should have a very similar set of common causes (potential confounders).

Substituting height for adiposity, I estimated associations with the availability of physical activity facilities and proximity to a fast-food outlet, using the same modelling approach as in the analyses where I examined associations between these same exposures and BMI, waist circumference and body fat. In these analyses, insofar as the assumption holds that the unobserved common causes of the exposure and the control are identical, a null association between the exposure and the control implies an absence of residual confounding of the exposure-outcome association.

3.4.5.2.Sensitivity to choice of outcome

In Chapter 4 I examine all three of BMI, WC and body fat as primary outcomes. In Chapter 6 I focus on BMI as the primary outcome, but I also perform the same analyses using WC and body fat. I do not do this is Chapter 5 because the genetic risk alleles used in that chapter are specifically linked to BMI. In Chapter 7 where I explore geographical

heterogeneity, I examine only BMI as an outcome, since results in earlier chapters were consistent across adiposity measures.

Throughout the thesis, I use a version of the BMI variable in which missing BMI values from the impedance machine are replaced with manual measurements taken by nurses at the assessment visit. This affects 1.4% of observations. In a sensitivity analysis in Chapter 4 I check whether results of the BMI models are robust to this imputation by running the same model excluding the observations relying on manually recorded BMI.

In Chapter 8, when I examine cancer-related hospital admissions, I also examine admissions related to breast and colorectal cancer specifically, as secondary outcomes.

3.4.5.3.Sensitivity to operationalisation of genetic risk

In Chapter 5 I construct weighted polygenic risk scores, which I examine as potential modifiers of the association between neighbourhood exposures and BMI. Although weighting of each SNP in the risk scores by its published effect size is standard practice[91], and appropriate due to the varying degree to which each SNP is associated with BMI, I perform sensitivity analyses using an unweighted version of each genetic risk score. In these, I expect to observe weaker evidence of a GxE interaction than in the main analysis, due to dilution of the effects of the more influential SNPs.

3.4.5.4. Sensitivity to covariate adjustment sets and missing data

In Chapter 4 I consider energy intake (in kJ, measured via 24-hour recall dietary assessment) as a possible confounder of the association between the physical activity environment and adiposity, and energy expenditure (self-reported) as a possible confounder of the association between the food environment and adiposity. As described in Chapter 4, adjustment for dietary intake risked introducing selection bias (due to missing data) and potentially also collider bias, hence the decision not to adjust for diet, and to instead examine the implications of that decision in sensitivity analyses.

In Chapter 5 I examine the possibility that the GxE interaction results might be biased by latent genetic structure in the UK Biobank sample, first by adjusting for 40 rather than the standard 10 first of the provided genetic ancestry principal components, and then by additionally adjusting for the geospatial coordinates (squared) of each individual's birth location.

In Chapter 8 I perform sensitivity analyses adjusting for an additional set of potential confounders that I had excluded from the primary analysis due to ambiguity regarding temporal precedence.

3.4.5.5.Sensitivity to sample restrictions

In Chapters 6 and 7 I exclude from the main analysis the 14% of UK Biobank participants living in non-urban postcodes. This is done because perceptions of proximity of food outlets and public amenity of green space are likely to differ in urban residential areas compared with non-urban areas[73–75] and I want to account for this potential source of heterogeneity in the sample. In Chapter 4 I simply adjusted for urban/non-urban status, but in Chapter 6 where combinations of exposures are considered and where public greenspace is also introduced as a possible effect modifier, I am more concerned with this issue and want to focus on drawing inferences for an urban population. To examine whether the findings differ by excluding non-urban participants rather than adjusting models of the full sample for urbanicity, I repeat the primary analyses on the full urban and non-urban sample combined, adjusting for urban/non-urban status.

In Chapter 5, the main analysis is restricted to the UK Biobank participants of White British ancestry because one of the two genetic risk scores is based only on SNPs associated with BMI in analyses of individuals with European ancestry. The other genetic risk score used in the main analyses includes SNPs associated with BMI in populations of non-European descent, so I undertook a sensitivity analysis that tested for GxE interactions with that risk score in a sample unrestricted by ethnicity, in order to assess generalisability of the primary findings to the broader population.

In Chapter 8, follow-up time in the main analysis starts immediately after baseline assessment (between 2007 and 2010). However, as some of the neighbourhood measures are based on secondary spatial data collected around 2012, I perform a sensitivity analysis in which I restrict follow up to 2012 onwards.

3.4.6. Ethics

Institutional ethical approval for this PhD was granted in September 2016 by the London School of Hygiene and Tropical Medicine's Observational/Intervention Research Ethics Committee (LSHTM Ethics Reference 11897). As a research project approved by UK Biobank's Principal Investigator, with oversight from the UK Biobank Ethics and Governance Council, Board and Access Sub-Committee, the project is also covered by ethics approval granted to UK Biobank by the North West Multi-centre Research Ethics

Committee (reference number 16/NW/0274). UK Biobank also secured the approval of the Patient Information Advisory Group (now National Information Governance Board for Health & Social Care) in England and Wales for gaining access to information that would allow it to invite people to participate, and the Community Health Index Advisory Group in Scotland.

REFERENCES

1 Greenland S, Rothman KJ, Lash T. Measures of effect and measures of association. In: Rothman KJ, Greenland S, Lash TL, eds. Modern Epidemiology, 3rd edn. Philadelphia, PA: Wolters Kluwer Health, 2008: 51–70.

2 VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiol Method 2014; 3: 33–72.

3 Kwan MP. The limits of the neighborhood effect: contextual uncertainties in geographic, environmental health, and social science research. Ann Am Assoc Geogr 2018; 108: 1482-1490.

4 Duncan DT, Goedel WC, Chunara R. Quantitative methods for measuring neighborhood characteristics in neighborhood health research. In: Duncan D, Kawachi I, editors. Neighborhoods and Health. 2nd ed. New York: Oxford University Press; 2018. p. 57–90.

5 Kwan MP. The uncertain geographic context problem. Ann Assoc Am Geogr 2012; 102: 958–68.

6 Lee C, Moudon AV. Correlates of walking for transportation or recreation purposes. J Phys Act Health 2006; 3: 77–98.

7 Grahn P, Stigsdotter UA. Landscape planning and stress. Urban For Urban Green 2003; 2: 1–18.

8 Coles RW, Bussey SC. Urban forest landscapes in the UK: progressing the social agenda. Landsc Urban Plan 2000; 52: 181–8.

9 Natural England. Nature Nearby - Accessible Natural Greenspace Guidance. 2010.

10 Chaix B, Merlo J, Evans D, Leal C, Havard S. Neighbourhoods in eco-epidemiologic research: delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. Soc Sci Med 2009; 69: 1306–10.

11 Perchoux C, Chaix B, Cummins S, Kestens Y. Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. Health Place 2013; 21: 86–93.

12 Kestens Y, Lebel A, Daniel M, Thériault M, Pampalon R. Using experienced activity spaces to measure foodscape exposure. Heal Place 2010; 16: 1094–103.

13 Hurvitz PM, Moudon AV. Home versus nonhome neighborhood: quantifying differences in exposure to the built environment. Am J Prev Med 2012; 42: 411–7.

14 Burgoine T, Monsivais P. Characterising food environment exposure at home, at work, and along commuting journeys using data on adults in the UK. Int J Behav Nutr Phys Act 2013; 10: 85.

15 Simpson JA, MacInnis RJ, Peeters A, Hopper JL, Giles GG, English DR. A comparison of adiposity measures as predictors of all-cause mortality: the Melbourne Collaborative Cohort Study. Obesity 2007; 15: 994–1003.

16 UK Biobank. UK Biobank Anthropometry. 2014. http://biobank.ctsu.ox.ac.uk/crystal/docs/Anthropometry.pdf.

17 Hu F. Obesity Epidemiology. Oxford: Oxford University Press, 2008.

18 Després J-P, Lemieux I. Abdominal obesity and metabolic syndrome. 2006; 444: 881–7.

19 Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006; 332: 1080.

20 The GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017; 377: 13–27.

21 Davillas A, Benzeval M. Alternative measures to BMI: exploring income-related inequalities in adiposity in Great Britain. Soc Sci Med 2016; 166: 223–32.

22 World Health Organization. International Classification of Diseases (ICD). 2011.

23 Leitzmann M, Powers H, Anderson AS, et al. European code against cancer 4th edition: physical activity and cancer. Cancer Epidemiol 2015; 39 Suppl 1: S46-55.

24 Norat T, Scoccianti C, Boutron-Ruault MC, et al. European code against cancer 4th edition: diet and cancer. Cancer Epidemiol 2015; 39 Suppl 1: S56–66.

25 Sallis JF, Glanz K. Physical activity and food environments: solutions to the obesity epidemic. Milbank Q 2009; 87: 123–54.

26 Diez Roux A-V. Neighborhoods and health: where are we and were do we go from here? Rev Epidemiol Sante Publique 2007; 55: 13–21.

27 Halonen JI, Stenholm S, Kivimäki M, et al. Is change in availability of sports facilities associated with change in physical activity? A prospective cohort study. Prev Med 2015; 73: 10–14.

28 Lee KY, Lee PH, Macfarlane D. Associations between moderate-to-vigorous physical activity and neighbourhood recreational facilities: the features of the facilities matter. Int J Environ Res Public Health 2014; 11: 12594–610.

29 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? Obes Rev 2011; 12: 501.

30 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. Br Med J 2014; 348: g1464.

31 Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. Obes Rev 2011; 12: e460-71.

32 Christian H, Giles-Corti B, Knuiman M, Timperio A, Foster S. The influence of the built environment, social environment and health behaviors on body mass index. results from RESIDE. Prev Med 2011; 53: 57–60.

33 Wilkins EL, Morris MA, Radley D, Griffiths C. Using Geographic Information Systems to measure retail food environments: discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). Health Place 2017; 44: 110–7.

34 Maguire ER, Burgoine T, Penney TL, Forouhi NG, Monsivais P. Does exposure to the food environment differ by socioeconomic position? Comparing area-based and person-centred metrics in the Fenland Study, UK. Int J Health Geogr 2017; 16: 1–14.

35 Sarkar C, Webster C, Gallacher J. Are exposures to ready-to-eat food environments associated with type 2 diabetes? A cross-sectional study of 347 551 UK Biobank adult participants. Lancet Planet Heal 2018; 2: e438–50.

36 Fraser LK, Edwards KL, Cade J, Clarke GP. The geography of fast food outlets: a review. Int J Environ Res Public Health 2010; 7: 2290–308.

37 Burgoine T, Alvanides S, Lake AA. Creating 'obesogenic realities'; do our methodological choices make a difference when measuring the food environment? Int J Health Geogr 2013; 12: 33.

38 Bivoltsis A, Cervigni E, Trapp G, Knuiman M, Hooper P, Ambrosini GL. Food environments and dietary intakes among adults: does the type of spatial exposure measurement matter ? A systematic review. Int J Health Geogr 2018; 17: 1–20.

39 Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. Annu Rev Public Health 2014; 35: 207–28.

40 Lachowycz K, Jones AP. Towards a better understanding of the relationship between greenspace and health: development of a theoretical framework. Landsc Urban Plan 2013; 118: 62–9.

41 Villanueva K, Badland H, Hooper P, et al. Developing indicators of public open space to promote health and wellbeing in communities. Appl Geogr 2015; 57: 112–9.

42 World Health Organization. Urban green spaces and health: a review of the evidence. 2016.

43 James P, Banay RF, Hart JE, Laden F. A review of the health benefits of greenness. Curr Epidemiol Reports 2015; 2: 131–42.

44 Alcock I, White MP, Wheeler BW, Fleming LE, Depledge MH. Longitudinal effects on mental health of moving to greener and less green urban areas. Environ Sci Technol 2014; 48: 1247–55.

45 Kaplan R. The nature of the view from home. Environ Behav 2001; 33: 507–42.

46 Weinstein N, Balmford A, DeHaan CR, Gladwell V, Bradbury RB, Amano T. Seeing community for the trees: the links among contact with natural environments, community cohesion, and crime. Bioscience 2015; 65: 1141–53.

47 Cox DTC, Shanahan DF, Hudson HL, et al. Doses of neighborhood nature: the benefits for mental health of living with nature. Bioscience 2017; 67: 147-155

48 Cox D, Shanahan D, Hudson H, et al. Doses of nearby nature simultaneously associated with multiple health benefits. Int J Environ Res Public Health 2017; 14: 172.

49 Mercille G, Richard L, Gauvin L, et al. The food environment and diet quality of urban-dwelling older women and men: assessing the moderating role of diet knowledge. Can J Public Heal 2016; 107: 34.

50 Stafford M, Cummins S, Macintyre S, Ellaway A, Marmot M. Gender differences in the associations between health and neighbourhood environment. Soc Sci Med 2005; 60: 1681–92.

51 Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. Epidemiol Rev 2009; 31: 7–20.

52 Jones-Smith JC, Karter AJ, Warton EM, et al. Obesity and the food environment: income and ethnicity differences among people with diabetes: the Diabetes Study of Northern California (DISTANCE). Diabetes Care 2013; 36: 2697–705.

53 Christine PJ, Auchincloss AH, Bertoni AG, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis (MESA). JAMA Intern Med 2015; 175: 1311–20.

54 Burgoine T, Sarkar C, Webster CJ, Monsivais P. Examining the interaction of fast-food outlet exposure and income on diet and obesity: evidence from 51,361 UK Biobank participants. Int J Behav Nutr Phys Act 2018; 15: 1–12.

55 Loos RJF. Recent progress in the genetics of common obesity. Br J Clin Pharmacol 2009; 68: 811–29.

56 Yeo GSH. The role of the *FTO* (Fat Mass and Obesity Related) locus in regulating body size and composition. Mol Cell Endocrinol 2014; 397: 34–41.

57 Waalen J. The genetics of human obesity. Transl Res 2014; 164: 293–301.

58 Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015; 518: 197–206.

59 Tyrrell J, Wood AR, Ames RM, et al. Gene-obesogenic environment interactions in the UK Biobank study. Int J Epidemiol 2017; 46: 559–75.

60 Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. Am J Hum Genet 2014; 95: 301–7.

61 Graff M, Richardson AS, Young KL, et al. The interaction between physical activity and obesity gene variants in association with BMI: does the obesogenic environment matter? Health Place 2016; 42: 159–65.

62 Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including *CADM2* and *APOE*. Int J Obes. 2018; 42: 1161–76.

63 Yen IH, Michael YL, Perdue L. Neighborhood environment in studies of health of older adults: a systematic review. Am J Prev Med 2009; 37: 455–63.

64 Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health 2001; 55: 111–22.

65 Arcaya M, Tucker-Seeley R, Kim R, Schnake-Mahl A, So M, Subramanian SV. Research on neighborhood effects on health in the United States: a systematic review of study characteristics. Soc Sci Med 2016; 168: 16–29.

66 Smith DM, Cummins S, Taylor M, et al. Neighbourhood food environment and area deprivation: spatial accessibility to grocery stores selling fresh fruit and vegetables in urban and rural settings. Int J Epidemiol 2010; 39: 277–84.

67 Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? Int J Behav Nutr Phys Act 2007; 4: 32

68 Barrington WE, Stafford M, Hamer M, Beresford SAA, Koepsell T, Steptoe A. Neighborhood socioeconomic deprivation, perceived neighborhood factors, and cortisol responses to induced stress among healthy adults. Heal Place 2014; 27: 120–6.

69 Hobbs M, Green MA, Griffiths C, et al. Access and quality of parks and associations with obesity: a cross-sectional study. SSM - Popul Heal 2017; 3: 722–9.

70 CABE. Urban Green Nation: Building the Evidence Base. London, UK, 2010.

71 UK Data Service. UK Data Service Census Deprivation data. 2012. http://census.ukdataservice.ac.uk/get-data/related/deprivation.aspx (accessed Sept 25, 2018).

72 Rae A. A Land Cover Atlas of the UK. 2017. DOI:10.15131/shef.data.5266495.

73 Maas J, Verheij RA, Groenewegen PP, De Vries S, Spreeuwenberg P. Green space, urbanity, and health: How strong is the relation? J Epidemiol Community Health 2006; 60: 587–92.

74 Coldwell DF, Evans KL. Contrasting effects of visiting urban green-space and the countryside on biodiversity knowledge and conservation support. PLoS One 2017; 12: e0174376.

75 Dennis M, James P. Evaluating the relative influence on population health of domestic gardens and green space along a rural-urban gradient. Landsc Urban Plan 2017; 157: 343–51.

76 Fornara F, Carrus G, Passafaro P, Bonnes M. Distinguishing the sources of normative influence on proenvironmental behaviors: The role of local norms in household waste recycling. Gr Process Intergr Relations 2011; 14: 623–35.

77 Patterson C, Hilton S. Normalisation and Stigmatisation of Obesity in UK Newspapers: a Visual Content Analysis. Open Obes J 2013; 5: 82–91.78 Scholes S, Pickering K, Rayat P, National Centre for Social Research. Neighbourhood Statistics, LA model-based estimates of healthy lifestyles

behaviours - 2003-05. 2007. https://digital.nhs.uk/data-and-information/publications/statistical/neighbourhood-statistics-mental-health-adults-accessing-nhs-specialist-mental-health-services/neighbourhood-statistics-la-model-based-estimates-of-healthy-lifestyles-behaviours-2003-05 (accessed July 26, 2018).

79 Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: The R package 'dagitty'. Int J Epidemiol 2016. DOI:10.1093/ije/dyw341.

80 Chaix B, Leal C, Evans D. Neighborhood-level Confounding in Epidemiologic Studies: Unavoidable Challenges, Uncertain Solutions. Epidemiology 2010; 21.

81 Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. Int J Epidemiol 2016; 45: 565-75.

82 Vanderweele T. Explanation in Causal Inference: Methods for Mediation and Interaction. New York : Oxford University Press, 2015.

83 Office for National Statistics. Regional gross disposable household income, UK 1997 to 2016. 2018.

84 Haworth S, Mitchell R, Corbin L, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. Nat Commun 2019; 10: 333.

85 Leckie G. Module 5: Introduction to Multilevel Modelling Stata Practical. Centre for Multilevel Modelling. 2010.

86 Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychol Methods 2007; 12: 121–38.

87 Li R, Chambless L. Test for additive interaction in proportional hazards models. Ann Epidemiol 2007; 17: 227–36.

88 Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012; 41: 514–20.

89 Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007; 4: e297.

90 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010; 21: 383–8.

91 Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet 2013; 9: e1003348.

Chapter 4. NEIGHBOURHOOD BUILT ENVIRONMENTS AND ADIPOSITY

4.1. Introduction

In this chapter I present an examination of the cross-sectional associations between characteristics of the neighbourhood built environment and three measures of adiposity described in the preceding chapter. The main part of this current chapter has now been published as a peer-reviewed paper in *The Lancet Public Health*, and is presented here in its published form.

The focus of the research paper is on the independent associations of neighbourhood food and physical activity environments with adiposity, along with possible effect modification by sex and household income. This paper sets up the foundation for subsequent chapters, by establishing the relationships within this cohort between adiposity and two of the central exposures on which I focus in this thesis. The research papers making up the three chapters following this one build on this by examining various forms of effect heterogeneity in these primary relationships – from interactions with genetic risk of obesity (Chapter 5), to modification of the effects of one neighbourhood characteristic by others (Chapter 6), through to geographical heterogeneity in the relationships across England (Chapter 7).

As additional material in this chapter, I also examine the association between neighbourhood greenspace and adiposity. In the final research paper (Chapter 8), I step from adiposity as the outcome, to more distal outcomes – namely CVD and cancer – and from a cross-sectional to longitudinal study design. In that final paper I explore an additional exposure (neighbourhood greenspace). For completeness I therefore include at the end of the current chapter an additional analysis of the association between the neighbourhood greenspace measure that I use in Chapter 8, and adiposity, mirroring the main analyses of the first research paper. This greenspace measure did not become available until after the research paper in the current chapter was published. Furthermore, the relationship between greenspace and adiposity is not a primary focus of this thesis because other researchers using UK Biobank have already published results of a similar analysis using a different greenspace measure available in UKBUMP.

4.2. Research Paper 1

Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank

Note: Further supplementary material that was published alongside, or referred to in, the following research paper is included in Appendix One.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1510923	**Title**	Ms
First Name(s)	Kate		
Surname/Family Name	Mason		
Thesis Title	Where and for whom does the neighbourhood built environment matter for obesity and health? Examining sources of effect heterogeneity at multiple scales in the UK adult population		
Primary Supervisor	Neil Pearce		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	The Lancet Public Health		
When was the work published?	12 December 2017		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	In collaboration with SC and with input from NP, I developed the research questions and designed the analysis. I then independently undertook the data management, statistical analysis, and writing of the first draft manuscript. SC and NP contributed to the interpretation of results and drafting of the manuscript, and I then produced the final version for publication.

SECTION E

Student Signature	[redacted]
Date	01/07/2019

Supervisor Signature	[redacted]
Date	01/07/2019

Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank

Kate E Mason, Neil Pearce, Steven Cummins

Summary

Background The built environment might be associated with development of obesity and related disorders. We examined whether neighbourhood exposure to fast-food outlets and physical activity facilities were associated with adiposity in UK adults.

Methods We used cross-sectional observational data from UK Biobank. Participants were aged 40–70 years and attended 21 assessment centres between 2006 and 2010. Using linked data on environments around each participant's residential address, we examined whether density of physical activity facilities and proximity to fast-food outlets were associated with waist circumference, body-mass index (BMI), and body fat percentage. We used multilevel linear regression models adjusted for potential confounders, and conducted several sensitivity analyses.

Findings Complete case sample sizes were 401 917 (waist circumference models), 401 435 (BMI), and 395 640 (body fat percentage). Greater density of physical activity facilities within 1000 m of home was independently associated with smaller waist circumference and lower BMI and body fat percentage. Compared with people with no nearby facilities, those with at least six facilities close to home had 1·22 cm smaller waist circumference (95% CI −1·64 to −0·80), 0·57 kg/m² lower BMI (−0·74 to −0·39), and 0·81 percentage points lower body fat (−1·03 to −0·59). Living further from a fast-food outlet was weakly associated with waist circumference and BMI, mostly among women. Compared with people living fewer than 500 m from a fast-food outlet, those living at least 2000 m away had 0·26 cm smaller waist circumference (−0·52 to 0·01).

Interpretation This study shows strong associations between high densities of physical activity facilities and lower adiposity for adults in mid-life. We observed weaker associations for access to fast food, but these are likely to be underestimated owing to limitations of the food environment measure. Policy makers should consider interventions aimed at tackling the obesogenic built environment.

Funding Commonwealth Scholarship Commission, Wellcome Trust Institutional Strategic Support Fund.

Lancet Public Health 2018; 3: e24–33

Published Online December 12, 2017 http://dx.doi.org/10.1016/S2468-2667(17)30212-8

See Comment page e4

Faculty of Epidemiology and Population Health (K E Mason MPH), **Department of Medical Statistics and Centre for Global Non-Communicable Diseases** (Prof N Pearce PhD), **and Department of Social & Environmental Health Research** (Prof S Cummins PhD), **London School of Hygiene & Tropical Medicine, London, UK**

Correspondence to: Ms Kate E Mason, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK **kate.mason@lshtm.ac.uk**

Introduction

Obesity is strongly linked to a range of chronic diseases, including type 2 diabetes and cardiovascular disease, and contributes substantially to excess morbidity, mortality, and rising health-care costs globally.[1,2] Across the world, increasing urbanisation is now recognised as a key driver of obesity and related non-communicable diseases, prompting calls to improve understanding of how urban environmental factors influence health.[3]

Particularly in urban areas, features of neighbourhood environments, such as access to unhealthy food and few opportunities for physical activity, might be associated with the development of obesity and related disorders. Collectively, such features are often referred to as the obesogenic environment,[4] and their presence and unequal distribution might partly explain rises in obesity prevalence and persistent social and geographical inequalities in obesity.[5] Although much research has been done on the influence of various neighbourhood features on obesity, consistent evidence remains elusive.[4] To further assess the potential of neighbourhoods to influence obesity outcomes via both energy intake and energy expenditure, and in particular the role of neighbourhood commercial resources, this Article focuses on two key features of the local residential environment: proximity to fast-food outlets and density of formal physical activity facilities. Recent studies indicate that about 21% of UK adults eat takeaway meals at home weekly[6] and about 18% regularly use gyms, with smaller proportions participating in other forms of physical activity likely to involve formal facilities (eg, swimming and team or racquet sports).[7]

Previous research suggests that access to fast-food outlets near the home might be a determinant of weight status and obesity,[8] although most supporting evidence comes from the USA, where the structure of the built environment differs from countries such as the UK. Although a recent study[9] in Cambridgeshire, UK, found that exposure to fast-food outlets near home was associated with both body-mass index (BMI) and odds of

Research in context

Evidence before this study

We searched databases including PubMed, Web of Science, and Google Scholar, up to Aug 31, 2017, using various combinations of search terms including "neighbourhood", "built environment", "obesity", "obesogenic", "adiposity", "food environment", and "physical activity", as well as hand-searching reference lists of relevant papers and tables of contents of recent or special editions of relevant journals. Our review of the literature was restricted to English language publications.

Much research has been done in recent decades into the effects of neighbourhood resources on obesity risk and related health behaviours. Features of the neighbourhood environment are hypothesised to affect diet quality through opportunities to consume healthy or unhealthy food (eg, access to fast food or grocery stores) and either encourage or discourage physical activity (eg, greenspace and street connectivity). However, the evidence base regarding so-called obesogenic environmental effects remains equivocal.

Many studies suggest that access to fast-food outlets in the local residential environment is a determinant of adiposity, but most evidence comes from the USA. Relatively little research has assessed the importance of local access to formal facilities for recreational physical activity (eg, gyms, leisure centres, swimming pools, and playing fields), instead focusing on broader urban design features such as walkability. In the UK, as elsewhere, many studies have been limited to a narrow geographical focus, relied on samples not sufficiently powered for subgroup analyses, or been prone to bias from residual confounding and exposure misclassification.

Added value of this study

This study is one of the first published to use UK Biobank to examine associations between features of the neighbourhood built environment and adiposity. The sample is made up of adults in mid-life—a crucial period of the lifecourse for the development of chronic disease. By using a very large dataset that covers much of the UK, we were able to provide evidence that relates to a wider geographical area than do most UK-based studies, and to examine sex and income differences. Making use of extensive covariate data available in UK Biobank, we were able to more comprehensively adjust for sources of confounding than have many other studies, and with additional sensitivity analyses, we were able to examine the robustness of our findings to residual confounding and different model specifications, further strengthening our findings.

We examined two features of the neighbourhood residential environment: proximity to fast-food outlets and density of formal physical activity facilities. We found evidence of a strong and graded inverse association between number of physical activity facilities close to home and three different adiposity measures. We observed a weaker association between the fast-food environment and adiposity, but limitations of the available food environment measure, such as misclassification of some fast-food outlets as restaurants and the inability to simultaneously account for both healthy and unhealthy food outlets, are likely to have attenuated the results. We also observed income and sex differences.

Implications of all the available evidence

The results of this study provide evidence to support the hypothesis that increasing access to local physical activity facilities and, possibly, reducing access to fast food close to residential areas has the potential to reduce overweight and obesity at the population level. Policy makers should consider interventions aimed at modifying residential environments to better facilitate healthy lifestyles, but recognising that such an approach might be more effective in some groups than in others. Future research should involve further interrogation of this rich data resource to examine geographical and other forms of heterogeneity in the effects of obesogenic environments on health to best target interventions.

obesity in adults, other international research has not consistently replicated findings from the USA.[10–12]

A range of neighbourhood attributes have also been found to potentially influence physical activity;[13] however, evidence of associations between obesity and neighbourhood attributes such as walkability and greenspace is mixed.[4,14] Relatively little research has assessed the importance of local access to formal facilities for recreational physical activity, such as gyms, swimming pools, and playing fields, although recent research in the UK indicates a possible relationship.[15]

Most studies in the UK have focused on particular geographical areas, such as individual cities or regions,[9,11,15] with only one UK-wide study,[16] which linked several neighbourhood-level contextual factors with obesity. Much like the international evidence, findings from the UK are mixed. This might be partly due to many studies being insufficiently powered, based on self-reported outcome measures, and focused on a narrow geographical area.

To improve understanding of how neighbourhood features influence adiposity and obesity-related outcomes in the UK, analyses of high-quality, individual, objective health data linked to environmental exposures measured at the individual address level are needed, for the whole of the UK. Using observational data from UK Biobank—a large sample of adults in a crucial period of the lifecourse for the development of chronic disease—we assessed whether the number of formal physical activity facilities near an individual's place of residence and proximity to fast-food outlets are independently associated with objectively measured adiposity. We also explore whether these associations differ by sex or income, and whether findings might be affected by residual confounding.

Methods

Study population

We used cross-sectional baseline data from UK Biobank (project 17380), a large, population-based cohort; the scientific rationale, study design, and survey methods for this project have been described elsewhere.[17] Data were potentially available from 502 656 individuals who had visited the 22 UK Biobank assessment centres across the UK between 2006 and 2010. Individuals aged 40–69 years living within a 25-mile radius of an assessment centre and listed on National Health Service patient registers were invited to participate in the UK Biobank study. The age range was chosen by UK Biobank as an important period for the development of many chronic diseases. The final recruited sample was aged 37–73 years, with more than 99% of participants aged 40–69 years.

Linked to UK Biobank is a high-resolution spatial database of objectively measured characteristics of the physical environment surrounding each participant's exact residential address, derived from multiple national spatial datasets.[18] The measures of the local environment include densities of various land uses, proximity to various health-relevant destinations (eg, general practitioners practices, industrial sites, fast-food outlets), greenspace, street-network accessibility, and pollution. The metrics were constructed using data collected in 2010, as close as possible to the baseline assessment of individuals.[18]

UK Biobank has ethics approval from the North West Multi-centre Research Ethics Committee (reference 16/NW/0274).

Exposures

We defined the physical activity environment as the density (count) of formal physical activity facilities within a 1000 m street-network buffer around each individual's place of residence. Formal physical activity facilities were defined at address level as any land use classified in the Commercial-Leisure subcategory of the UK Ordnance Survey AddressBase Premium database.[18] This subcategory comprises a range of indoor and outdoor facilities designed for sporting and leisure activities, such as gyms, swimming pools, and playing fields (see appendix for details). We did not include informal physical activity facilities, such as public parks and cycling paths (except where covered by the above classification—eg, playing fields), because formal physical activity facilities are understudied as a neighbourhood health resource and might have different drivers of use. Because many of these facilities are commercial, they are also potentially modifiable via regulatory and commercial levers that are less relevant to the informal physical activity environment.

For each individual, the street-network distance (in metres) from residential address to the nearest fast-food outlet, classified as "hot/cold fast-food outlet/takeaway" in the UK Ordnance Survey AddressBase Premium database,[18] was available. We used these distances and the distribution of data to categorise individuals as living closer than 500 m, 500–999 m, 1000–1999 m, or at least 2000 m from their nearest fast-food outlet. A proximity measure was used by contrast with the physical activity environment density measure because no equivalent density measure was available for the food environment. Density and proximity measures in the UK have been shown to be correlated despite being theoretically distinct.[19]

Adiposity measures

We used three adiposity measures: waist circumference, BMI (calculated from height and weight), and body fat percentage (measured by bioimpedance). Measurements were made by trained staff using standard procedures.[17] We centred all three metrics around their mean and treated them as continuous variables.

Potential confounders and model adjustments

We identified potential confounders for each of the two environmental exposures on the basis of the existing literature, and summarised this information in directed acyclic graphs to assess which should be adjusted for in the main analyses (appendix). Potential confounders included individual demographic and socioeconomic variables and several local area characteristics. Because the fast-food and physical activity environments are also associated with one another, features of the food environment might confound associations between the physical activity environment and adiposity, and the physical activity environment might confound associations between the food environment and adiposity. Although individual diet and physical activity behaviours are associated with adiposity, physical activity is unlikely to be a common ancestor of the food environment and adiposity, and similarly diet will not predict the physical activity environment. Furthermore, conditioning on these behaviours risks inducing collider bias[20] by opening backdoor pathways through genetic risk, prior behaviours, and prior adiposity.

On the basis of these causal diagrams, we adjusted the final models for age (years), sex (male or female), ethnicity (white, south Asian, black, other Asian, mixed white and black, mixed white and Asian, mixed other, or other), highest education level attained (college or university degree; A levels, AS levels, or equivalent [academic advanced levels, post compulsory education]; O levels, GCSEs, or equivalent [higher secondary education]; CSE or equivalent [secondary education]; National Vocational Qualification, Higher National Diploma, Higher National Certificate, or equivalent [vocational qualifications]; other professional qualification; or none of the above), annual household income (<£18 000, £18 000–30 999, £31 000–51 999, £52 000–100 000, or >£100 000), employment status (paid work, retired, unable to work, unemployed, or other),

area deprivation (Townsend score), urbanicity (urban or non-urban), and neighbourhood residential density (count of residential dwellings within a 1000 m street-network buffer of home address, log transformed). We adjusted models of the physical activity environment and adiposity for fast-food proximity, and adjusted models of the fast-food environment and adiposity for density of physical activity facilities.

We used negative control analyses to analyse possible residual confounding. Negative control analysis is a technique that can help to detect residual confounding in observational studies.[21] Because studies of neighbourhoods and health are potentially susceptible to residual confounding by factors such as other neighbourhood characteristics and residential segregation, we employed this technique by doing analyses similar to our primary analyses, but in which the outcome was a variable not expected to be associated with our exposures; specifically, we used height (cm) as a negative control outcome because height is unlikely to be related to neighbourhood environment in adulthood, but, as with adiposity, it is correlated with various sociodemographic characteristics of individuals. If the main associations for adiposity were residually confounded, we would expect to observe a spurious association between the environmental exposures and height.

Statistical analysis

We used multilevel, multiple linear regression models with random intercepts and random coefficients for the main exposure to estimate independent associations between each environmental exposure and each adiposity outcome, accounting for the nesting of individuals within assessment centres. We initially adjusted only for age and sex (model 0), then for likely demographic confounders (age, sex, ethnicity, area deprivation, and urbanicity; model 1), then further adjusted for individual-level socioeconomic characteristics (income, education, and employment status; model 2) and, finally, for the non-exposure environmental feature (proximity to fast food or density of physical activity facilities) and neighbourhood residential density (model 3). As well as adjusting for potential confounding by sex and income, we also tested fully adjusted models for effect modification by these variables. We report stratified results where models with interaction terms for sex or income were statistically different from those without (likelihood ratio test $p<0·05$). Finally, we estimated the same models using height as a negative control outcome.

We designed sensitivity analyses to test the robustness of our findings to different model specifications. These analyses examined the effect of further adjustment for the behavioural variables diet and physical activity; the extent to which any such effect was driven by selection bias due to missing data; and the sensitivity of our models to the choice of BMI measure.

We did all analyses in Stata SE version 14.2.

Role of the funding source

The funder had no role in the design, conduct, or writing up of this study. The corresponding author had full access to all of the data and final responsibility to submit for publication.

	Data (n=498 822)
Adiposity	
Waist circumference, cm	90·3 (13·5)
Range	20 to 197
Data missing	2110 (0·4%)
Body-mass index, kg/m²	27·4 (4·8)
Range	13 to 68
Data missing	10 014 (2·0%)
Body fat	31·5% (8·5)
Range	5% to 70%
Data missing	10 286 (2·1%)
Environment	
Physical activity environment	
Number of facilities in 1000 m buffer	1 (0 to 3)
0	150 211 (31·2%)
1	96 031 (20·0%)
2 to 3	114 693 (23·8%)
4 to 5	58 217 (12·1%)
≥6	61 978 (12·9%)
Range	0 to 39
Data missing	17 692 (3·5%)
Fast-food environment	
Distance to nearest outlet, m	1136 (615 to 2197)
<500 m	88 804 (18·5%)
500 m to 999 m	124 698 (25·9%)
1000 m to 1999 m	133 401 (27·7%)
≥2000 m	134 180 (27·9%)
Range	0 to 96 538
Data missing	17 719 (3·6%)
Covariates	
Age, years*	56·5 (8·1)
Range	40 to 70
Data missing	15 (0·0%)
Sex	
Female	271 384 (54·4%)
Male	227 438 (45·6%)
Data missing	1 (0·0%)
Ethnicity	
White	469 209 (94·6%)
South Asian or south Asian British	8015 (1·6%)
Black or black British	8038 (1·6%)
Chinese or other (non-South) Asian	3367 (0·7%)
Mixed: white and black	1029 (0·2%)
Mixed: white and Asian	827 (0·2%)
Mixed: detail unknown	1073 (0·2%)
Other	4521 (0·9%)
Data missing	2743 (0·5%)

(Table 1 continues on next page)

	Data (n=498 822)
(Continued from previous page)	
Income	
Less than £18 000	97 221 (22·9%)
£18 000–30 999	108 197 (25·4%)
£31 000–51 999	110 790 (26·0%)
£52 000–100 000	86 280 (20·3%)
Greater than £100 000	22 933 (5·4%)
Data missing	73 401 (14·7%)
Education†	
College or university degree	161 200 (32·7%)
A levels, AS levels, or equivalent	55 331 (11·2%)
O levels, GCSEs, or equivalent	105 218 (21·4%)
CSE or equivalent	26 893 (5·5%)
NVQ, HND, HNC, or equivalent	32 734 (6·6%)
Other professional qualifications	25 810 (5·2%)
None of the above	85 291 (17·3%)
Data missing	6345 (1·3%)
Employment status	
Paid employment or self-employed	284 873 (57·4%)
Retired	165 977 (33·5%)
Unable to work	16 654 (3·4%)
Unemployed	8225 (1·7%)
Home duties, carer, student, volunteer, or other	20 171 (4·1%)
Data missing	2922 (0·6%)
Area deprivation (Townsend index)	−2·1 (−3·6 to 0·5)
Data missing	626 (0·1%)
Urbanicity	
Urban	425 082 (86·1%)
Non-urban	68 665 (13·9%)
Data missing	5075 (1·0%)
Residential density‡	1899 (1095 to 3102)
Range	1 to 22 306
Data missing	17 705 (3·5%)
Total dietary energy intake, kJ	8752·6 (2784·6)
Range	0 to 19 995
Data missing	288 690 (57·9%)
Physical activity (MET min per week)	1666 (743 to 3413)
Range	0 to 32 130
Data missing	43 656 (8·8%)
Assessment area	
Manchester	13 940 (2·8%)
Oxford	14 062 (2·8%)
Cardiff	17 882 (3·6%)
Glasgow	18 651 (3·7%)
Edinburgh	17 201 (3·4%)
Stoke	19 440 (3·9%)
Reading	29 417 (5·9%)
Bury	28 335 (5·7%)
Newcastle	37 008 (7·4%)
Leeds	44 209 (8·9%)
Bristol	43 015 (8·6%)
Central London	12 583 (2·5%)
Nottingham	33 877 (6·8%)
Sheffield	30 397 (6·1%)
Liverpool	32 818 (6·6%)
Middlesbrough	21 289 (4·3%)
Hounslow	28 879 (5·8%)
Croydon	27 385 (5·5%)
Birmingham	25 503 (5·1%)
Swansea	2281 (0·5%)
Wrexham	649 (0·1%)

Summary statistics were examined for the full sample of participants from the 21 areas linked to the environmental dataset. Data are n (%), mean (SD), or median (IQR), and are given for the complete case sample for each category, with number and percentage of records with missing data displayed for all variables with missing entries. Due to rounding error, some percentages sum to more than 100%. MET=metabolic equivalent of task. *Our analytical sample included people aged 70 years at the time of assessment, but excluded nine individuals with complete data who were younger than 40 years or older than 70 years. †UK qualifications break down into academic advanced levels (A levels, AS levels, or equivalent, post compulsory education), higher secondary education (O levels, GCSEs, or equivalent), secondary education (CSE or equivalent), and vocational qualifications (NVQ, HND, HNC, or equivalent). ‡Residential address points per 1000 m buffer.

Table 1: **Characteristics of study participants**

Results

Environmental data were available for individuals from 21 of the 22 assessment areas; no environmental data were collected for participants assessed in Stockport, where the pilot study was conducted. Of the primary covariates (excluding diet and physical activity), income had the most missing data, followed by residential density, BMI, and body fat percentage (table 1). All other covariates were missing at a frequency of approximately 1% or less (table 1). After the exclusion of cases missing data on key covariates, the final complete case sample sizes were 401 917 for waist circumference analyses, 401 435 for BMI analyses, and 395 640 for body fat analyses. Excluded observations were very similar to complete cases in terms of neighbourhood exposures and adiposity, but were more often people from more deprived postcodes, less highly educated, more likely to be retired, and more likely to be of south Asian or black ethnicity.

The mean waist circumference of the full sample with environmental data available was 90·3 cm, mean BMI was 27·4 kg/m², and mean body fat percentage was 31·5% (table 1). The median number of formal physical activity facilities within a 1000 m street-network distance of participants' homes was one, with a third of participants having no facilities close to home (table 1). Participants lived a median of 1136 m from a fast-food outlet, with nearly a fifth living within 500 m of such an outlet (table 1).

In fully adjusted models, adiposity was lower among people with greater access to local physical activity facilities compared with those with fewer facilities near

	Model 0	Model 1	Model 2	Model 3
Waist circumference, cm (n=401 917)				
Number of facilities				
0	0 (ref)	0 (ref)	0 (ref)	0 (ref)
1	0·15 (−0·02 to 0·32)	−0·15 (−0·30 to −0·01)	−0·13 (−0·26 to 0·00)	−0·19 (−0·32 to −0·06)
2–3	0·08 (−0·20 to 0·35)	−0·43 (−0·70 to −0·17)	−0·29 (−0·51 to −0·08)	−0·40 (−0·62 to −0·17)
4–5	−0·14 (−0·55 to 0·27)	−0·80 (−1·19 to −0·42)	−0·51 (−0·82 to −0·19)	−0·65 (−0·96 to −0·33)
≥6	−0·67 (−1·25 to −0·09)	−1·51 (−2·04 to −0·98)	−1·03 (−1·45 to −0·62)	−1·22 (−1·64 to −0·80)
Body-mass index, kg/m² (n=401 435)				
Number of facilities				
0	0 (ref)	0 (ref)	0 (ref)	0 (ref)
1	0·04 (−0·04 to 0·12)	−0·08 (−0·15 to 0·00)	−0·06 (−0·13 to 0·01)	−0·07 (−0·14 to 0·00)
2–3	−0·03 (−0·15 to 0·10)	−0·22 (−0·34 to −0·09)	−0·16 (−0·26 to −0·05)	−0·18 (−0·28 to −0·08)
4–5	−0·18 (−0·36 to 0·00)	−0·43 (−0·59 to −0·26)	−0·30 (−0·43 to −0·17)	−0·33 (−0·46 to −0·20)
≥6	−0·42 (−0·67 to −0·17)	−0·73 (−0·96 to −0·50)	−0·52 (−0·69 to −0·34)	−0·57 (−0·74 to −0·39)
Body fat, % (n=395 640)				
Number of facilities				
0	0 (ref)	0 (ref)	0 (ref)	0 (ref)
1	0·03 (−0·07 to 0·14)	−0·11 (−0·20 to −0·01)	−0·08 (−0·16 to 0·00)	−0·11 (−0·20 to −0·03)
2–3	−0·07 (−0·24 to 0·11)	−0·30 (−0·46 to −0·13)	−0·21 (−0·34 to −0·07)	−0·27 (−0·40 to −0·13)
4–5	−0·28 (−0·53 to −0·02)	−0·58 (−0·81 to −0·35)	−0·40 (−0·58 to −0·22)	−0·48 (−0·67 to −0·29)
≥6	−0·63 (−0·96 to −0·30)	−1·00 (−1·29 to −0·70)	−0·71 (−0·92 to −0·49)	−0·81 (−1·03 to −0·59)

Density is defined as number of physical activity facilities in a 1000 m street-network buffer. Data are mean difference (95% CI). Model 0: adjusted for age and sex. Model 1: model 0 plus adjustment for ethnicity, urban or non-urban status, and area deprivation. Model 2: model 1 plus adjustment for individual socioeconomic characteristics (income, education, and employment status). Model 3: model 2 plus adjustment for residential density and distance to nearest fast-food outlet.

Table 2: **Associations between density of physical activity facilities and adiposity outcomes: multilevel regression results**

home (table 2). Compared with the reference category (no nearby physical activity facilities), the waist circumference of those with at least six facilities nearby was, on average, 1·22 cm smaller (95% CI −1·64 to −0·80; p<0·0001); their BMI was 0·57 kg/m² lower (−0·74 to −0·39; p<0·0001); and their body fat was 0·81 percentage points lower (−1·03 to −0·59; p<0·0001). Regression coefficients decreased monotonically across categories of increasing density.

Main effect estimates were smallest in models adjusted only for age and sex (model 0; table 2). In fully adjusted models (model 3), physical activity environment coefficients were attenuated compared with models that also controlled for ethnicity, urbanicity, and area deprivation (model 1), but were larger in magnitude than intermediate models further adjusted for individual socioeconomic characteristics (model 2; table 2).

For the associations between physical activity facilities and adiposity, we found evidence of effect modification by sex (waist circumference p<0·0001; BMI p=0·0009; body fat p=0·0029) and income (waist circumference and BMI both p<0·0001; body fat p=0·0026), with an inverse association for all subgroups but stronger among women (figure 1) and people from higher-income households (figure 2).

We also estimated lower mean waist circumference with each categorical increase in distance to the nearest fast-food outlet, independent of the influence of the physical activity environment and all other covariates (ie, model 3; table 3). However, the only significant coefficients estimated were for intermediate proximity categories, and these were small: compared with participants living within 500 m of a fast-food outlet, the mean waist circumference among those living within 500–999 m was 0·15 cm smaller (95% CI −0·30 to −0·01, p=0·040) and 0·22 cm smaller (−0·44 to 0·00, p=0·049) for those living within 500–1499 m (table 3). For those living at least 2000 m from an outlet, the average decrease was 0·26 cm (95% CI −0·52 to 0·01; p=0·057). We observed a similar pattern of association for BMI. For body fat percentage, evidence of an association was weaker—eg, for people living at least 2000 m from a fast-food outlet, body fat was 0·10 percentage points lower (p=0·18). Sex modified these relationships for all three outcome measures (p<0·0001 for all measures; figure 1), whereas income did not (waist circumference p=0·094; BMI p=0·423; body fat p=0·083; appendix). With regard to sex, only among women did we observe an inverse dose–response association, with lower average adiposity across all measures the further a woman lived from a fast-food outlet, although only the waist circumference coefficient for the greatest distance category (≥2000 m) reached significance at the 5% level (0·36 cm decrease, 95% CI −0·72 to −0·01). We observed no such association for men.

Fast-food-environment coefficients in fully adjusted models were substantially attenuated compared with models adjusted only for age and sex, but were larger in magnitude than in both intermediate models, suggesting that failure to adjust for neighbourhood-level confounders attenuates results.

The negative control analysis, using height as an outcome, yielded the expected null results for the fast-food environment, but showed evidence of increasing height with an increasing number of formal physical activity facilities; regression coefficients were small in magnitude (all <0·52 cm difference) but significant for the subgroups of 2–3 facilities, 4–5 facilities, and six facilities or more (table 4), suggesting some possible residual confounding.

Additional adjustment for behavioural confounders had negligible impact on the regression coefficients

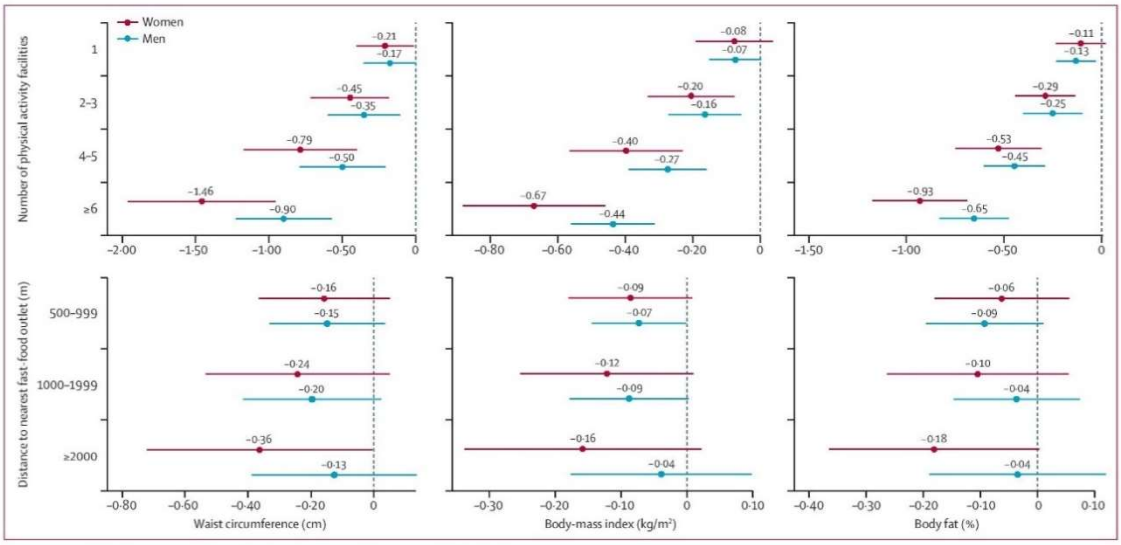


Figure 1: **Association between neighbourhood environmental features and adiposity, by sex**
Figure shows sex-stratified, fully adjusted mean differences in adiposity and associated 95% CIs. The dashed line at zero represents the reference category (no physical activity facilities with 1000 m of home or <500 m to nearest fast-food outlet).

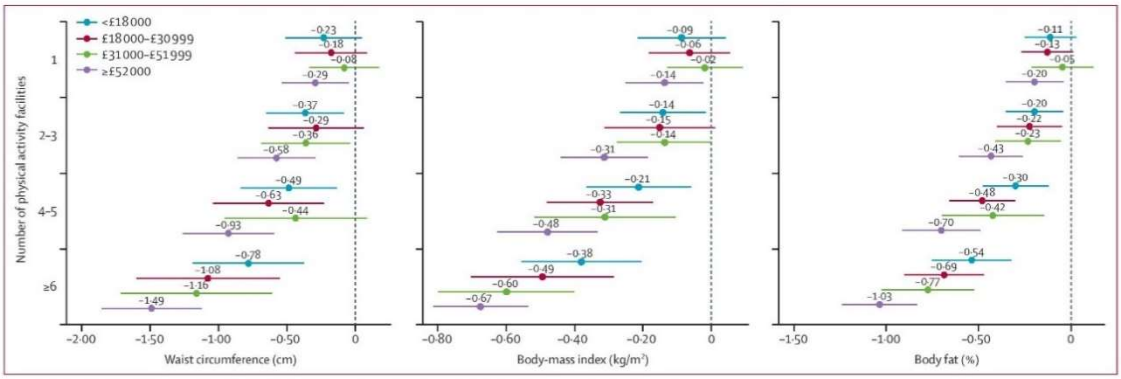


Figure 2: **Association between number of formal physical activity facilities and adiposity, by annual household income**
Figure shows annual-household-income-stratified, fully adjusted mean differences in adiposity and associated 95% CIs. The dashed line at zero represents the reference category (no physical activity facilities with 1000 m of home). Income-stratified results for the association between distance to nearest fast-food outlet and adiposity can be found in the appendix.

(appendix). Adjustment for diet in a sensitivity analysis of physical activity models did not lead to substantively different conclusions but did inflate some point estimates; however, sensitivity analyses that estimated model 3 on a sample restricted to people with dietary data produced almost identical results, indicating that the observed inflation is driven by sample restriction due to missing data rather than solely additional adjustment for diet (appendix). Models that excluded the small proportion of BMI values measured manually rather than with the bioimpedance machine yielded substantively identical results to those from the primary analysis (appendix).

Discussion

In this uniquely large and geographically diverse sample of adults in mid-life, we found consistent evidence that local access to formal physical activity facilities such as leisure centres, gyms, and sports fields is independently associated with adiposity. As the density of formal

	Model 0	Model 1	Model 2	Model 3
Waist circumference, cm (n=401 917)				
Distance				
<500 m	0 (ref)	0 (ref)	0 (ref)	0 (ref)
500–999 m	−0·48 (−0·67 to −0·29)	−0·08 (−0·24 to 0·07)	−0·07 (−0·22 to 0·07)	−0·15 (−0·30 to −0·01)
1000–1999 m	−0·69 (−1·01 to −0·38)	−0·04 (−0·29 to 0·22)	−0·08 (−0·27 to 0·12)	−0·22 (−0·44 to 0·00)
≥2000 m	−1·20 (−1·59 to −0·81)	−0·05 (−0·40 to 0·30)	−0·11 (−0·36 to 0·14)	−0·26 (−0·52 to 0·01)
Body-mass index, kg/m² (n=401 435)				
Distance				
<500 m	0 (ref)	0 (ref)	0 (ref)	0 (ref)
500–999 m	−0·20 (−0·29 to −0·11)	−0·04 (−0·10 to 0·03)	−0·03 (−0·09 to 0·03)	−0·08 (−0·14 to −0·02)
1000–1999 m	−0·24 (−0·39 to −0·09)	0·01 (−0·11 to 0·12)	−0·01 (−0·10 to 0·08)	−0·10 (−0·20 to −0·01)
≥2000 m	−0·40 (−0·60 to −0·21)	0·03 (−0·15 to 0·22)	0·01 (−0·13 to 0·14)	−0·10 (−0·24 to 0·04)
Body fat, % (n=395 640)				
Distance				
<500 m	0 (ref)	0 (ref)	0 (ref)	0 (ref)
500–999 m	−0·18 (−0·27 to −0·08)	−0·02 (−0·10 to 0·07)	−0·02 (−0·10 to 0·06)	−0·08 (−0·16 to 0·00)
1000–1999 m	−0·19 (−0·35 to −0·03)	0·07 (−0·07 to 0·21)	0·04 (−0·07 to 0·15)	−0·07 (−0·19 to 0·05)
≥2000 m	−0·46 (−0·67 to −0·24)	0·05 (−0·15 to 0·26)	0·00 (−0·14 to 0·15)	−0·10 (−0·26 to 0·05)

Distances given are distance to nearest fast-food outlet. Data are mean difference (95% CI). Model 0: adjusted for age and sex. Model 1: model 0 plus adjustment for ethnicity, urban or non-urban status, and area deprivation. Model 2: model 1 plus adjustment for individual socioeconomic characteristics (income, education, and employment status). Model 3: model 2 plus adjustment for residential density and density of local physical activity facilities.

Table 3: **Associations between proximity to fast-food outlets and adiposity outcomes: multilevel regression results**

	Height, cm (n=401 676)
Number of physical activity facilities in 1000 m street-network buffer*	
0	1 (ref)
1	0·09 (0·01 to 0·16)
2–3	0·21 (0·10 to 0·32)
4–5	0·38 (0·27 to 0·49)
≥6	0·51 (0·36 to 0·66)
Distance to nearest fast-food outlet†	
<500 m	1 (ref)
500–999 m	0·01 (−0·06 to 0·09)
1000–1999 m	−0·02 (−0·11 to 0·08)
≥2000 m	0·01 (−0·10 to 0·13)

Adjusted for age, sex, ethnicity, urban or non-urban status, area deprivation, income, education, employment status, and residential density. *Also adjusted for distance to nearest fast-food outlet. †Also adjusted for density of local physical activity facilities.

Table 4: **Results from negative control analyses**

physical activity facilities increased, waist circumference, BMI, and body fat percentage all decreased. We observed a similar but much weaker association between proximity to fast-food outlets and adiposity. This dataset provided a unique opportunity to examine these associations in a sample spanning diverse areas of the UK rather than being limited to a single study site.

Physical activity facilities in the local environment provide convenient opportunities for recreational physical activity. If improved access increases physical activity behaviour, we would expect to see a causal effect on adiposity. In income-stratified models, the inverse association we observed for all income groups was most marked among higher-income households. This finding is unsurprising given many facilities have costs attached to use, and has implications for municipal and private providers of physical activity facilities, who should be encouraged to invest in facilities in or near residential areas, but also to ensure that costs of access are managed to avoid inadvertently widening health inequalities.

Although our findings regarding physical activity facilities are consistent with a recent Scottish study[15] that found some measures of accessibility of physical activity facilities to be associated with BMI, our food environment findings are somewhat inconsistent with a recent study[9] in England that found a strong, graded association between fast-food access and BMI, whereas we observed only a weak association. Food outlet classification in the source database for the current study is supplied by local authorities and might include misclassification of some outlets as restaurants rather than fast-food outlets, potentially biasing our regression coefficients towards the null. Given the limits of the available data, we were also unable to account for other dimensions of the food environment, which might have biased our estimates (see appendix p 2). Indeed, empirical evidence from other studies suggests that simultaneously accounting for healthy and unhealthy food outlets yields larger and more precise estimates of health effects than when considering only a single dimension of the food environment.[9,22,23] This possibility of bias is further supported indirectly by the observation from our intermediate models that main effects were attenuated when no adjustment was made for other area-level confounders. There might also be local variability in the accuracy and completeness of these data, which we have not been able to assess. Stronger measures of the food environment linked to this dataset would provide more robust evidence on which to make more reliable inferences about the role of the fast-food environment in the UK. For many people, environmental determinants of diet and weight are also likely to include commuting routes and workplaces,[24] but we were unable to assess these. Associations between adiposity and neighbourhood food environments might also not be consistent across geographical regions, even within the same country.[25,26]

We found strong evidence of effect modification by sex, and stratified models showed modestly larger estimates of effects of both neighbourhood exposures on women's adiposity than on men's. Other studies have also observed

sex differences in neighbourhood effects, in which women in some age groups appeared more sensitive to health impacts of local environments.[27,28] These differences might relate to traditional gender roles that result in women spending more time than men in their local neighbourhood, although this has not been clearly established.[28]

The negative control analysis for the physical activity environment indicates possible residual confounding, because we observed an unexpected association with height. Comparison of standardised coefficients across the relevant models (not shown) suggests this residual confounding would only partially account for the observed effect. For the physical activity environment, we observed in intermediate models that failure to adjust for individual socioeconomic characteristics biased coefficients away from the null. Therefore, if individual socioeconomic position has not been adequately controlled for (eg, insufficient specificity in the categorical, non-equivalised income variable), main effects might be overestimated.

This study has several strengths. The unique scale and size of UK Biobank reflect a geographical coverage that no other similar studies have, particularly in the UK, and enables examination of subgroup heterogeneity often not possible in smaller studies. We investigated two neighbourhood exposures, based on exact home address, thought to influence the same outcome via separate pathways, and assessed their independent associations with adiposity by controlling one for the other. Both exposures are primarily commercial in nature, and therefore amenable to regulatory and market-based interventions. One of these exposures—formal physical activity facilities—has received relatively little research attention to date. We examined associations with multiple objective biomarkers of adiposity because available measures vary in how well they predict different health outcomes in different populations.[29] The consistency we observed across multiple outcome measures, especially for the physical activity environment, can be seen as providing stronger evidence of a potentially causal relationship. Sensitivity analyses suggest our main findings are robust to model misspecification, and that adjustment for health behaviours is not necessary, and in some cases might induce bias. Our findings were also consistent in preliminary analyses using alternative cutpoints for categories of the exposure variables (not shown). Although modest, the adiposity differences observed are averages across the sample, including people unlikely to eat fast food or use physical activity facilities regardless of local accessibility. If these observed differences do represent a causal relationship, the actual magnitude of effect would be larger among those likely to be affected.

In addition to weaknesses of the food environment measure highlighted above, the study has some further limitations. Although the UK Biobank sample is very large, the response rate was low (5·5%) and the sample shows evidence of so-called healthy volunteer bias.[30] The non-null associations we observed between access to physical activity facilities and height in a negative control analysis suggest that some unmeasured confounding of the physical activity environment associations with adiposity remained. Studies of neighbourhood effects are particularly susceptible to bias arising from residential mobility, where movement between neighbourhoods over time increases the risk of exposure misclassification, and leaves open the potential for reverse causation in cross-sectional analyses if, for example, individuals with lower adiposity choose to live in areas with more physical activity facilities. Current adiposity might also reflect exposure to neighbourhood environments earlier in life, posing a further challenge for causal inference. Therefore, an important strength of this sample is its stability: more than 65% of respondents had been resident at their current address for at least 10 years (mean 17·3 years [SD 11·8]). Despite this, some risk of exposure misclassification might remain owing to retailer turnover, especially for fast-food outlets. Furthermore, the relationship between the fast-food environment and adiposity is likely to be bidirectional, but within this study we were unable to disentangle the possible effects of retailers positioning fast-food outlets in areas of higher demand. Another limitation of our analyses is the unavailability of equivalent metrics for the physical activity and food environments that might allow a more direct comparison of their influence.

Research on neighbourhood environments and obesity has produced inconsistent findings to date. Our findings from a large and geographically diverse sample of adults in mid-life add support to the hypothesis that increasing access to physical activity facilities and, possibly, reducing access to fast food close to residential areas has the potential to reduce the prevalence of obesity and overweight at the population level, but highlight that this approach might be more effective for some groups than for others.

Contributors

KEM and SC designed the analysis with input from NP. KEM undertook the data management, statistical analysis, and writing of the manuscript. SC and NP contributed to the interpretation of results and drafting of the manuscript. All authors approved the final submitted version.

Declaration of interests

We declare no competing interests.

Acknowledgments

KM is funded by a Commonwealth Scholarship Commission PhD Scholarship. NP is supported by the Wellcome Trust Institutional Strategic Support Fund, 097834/Z/11/B through the Centre for Global NCDs. This research has been conducted using the UK Biobank Resource under Application Number 17380.

References

1 Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; **384:** 766–81.

2 Dobbs R, Sawers C, Thompson F, et al. Overcoming obesity: an initial economic analysis. London: McKinsey & Company, 2014.

3 The Lancet Diabetes & Endocrinology. Urbanisation, inequality, and non-communicable disease risk. *Lancet Diabetes Endocrinol* 2017; **5:** 313.

4 Mackenbach JD, Rutter H, Compernolle S, et al. Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. *BMC Public Health* 2014; **14:** 233.

5 Black JL, Macinko J. Neighborhoods and obesity. *Nutr Rev* 2008; **66:** 2–20.

6 Adams J, Goffe L, Brown T, et al. Frequency and socio-demographic correlates of eating meals out and take-away meals at home: cross-sectional analysis of the UK national diet and nutrition survey, waves 1–4 (2008–12). *Int J Behav Nutr Phys Act* 2015; **12:** 51.

7 Stamatakis E, Chaudhury M. Temporal trends in adults' sports participation patterns in England between 1997 and 2006: the Health Survey for England. *Br J Sports Med* 2008; **42:** 901–08.

8 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev* 2011; **12:** e95–106.

9 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 2014; **348:** g1464.

10 Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. *Obes Rev* 2011; **12:** e460–71.

11 Griffiths C, Frearson A, Taylor A, Radley D, Cooke C. A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK. *Int J Behav Nutr Phys Act* 2014; **11:** 138.

12 Christian H, Giles-Corti B, Knuiman M, Timperio A, Foster S. The influence of the built environment, social environment and health behaviors on body mass index. results from RESIDE. *Prev Med* 2011; **53:** 57–60.

13 Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the Built environment for physical activity. State of the science. *Am J Prev Med* 2009; **36:** S99–123.

14 Koohsari MJ, Mavoa S, Villanueva K, et al. Public open space, physical activity, urban design and public health: Concepts, methods and research agenda. *Health Place* 2015; **33:** 75–82.

15 Ellaway A, Lamb KE, Ferguson NS, Ogilvie D. Associations between access to recreational physical activity facilities and body mass index in Scottish adults. *BMC Public Health* 2016; **16:** 756.

16 Stafford M, Cummins S, Ellaway A, Sacker A, Wiggins RD, Macintyre S. Pathways to obesity: identifying local, modifiable determinants of physical activity and diet. *Soc Sci Med* 2007; **65:** 1882–97.

17 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007. http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf (accessed Nov 1, 2017).

18 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP)—a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21:** 135–48.

19 Burgoine T, Alvanides S, Lake AA. Creating 'obesogenic realities'; do our methodological choices make a difference when measuring the food environment? *Int J Health Geogr* 2013; **12:** 33.

20 Glymour MM, Greenland S. Causal Diagrams. In: Rothman KJ, Greenland S, Lash TL, eds. Modern epidemiology, 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008: 183–209.

21 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010; **21:** 383–88.

22 Clary CM, Ramos Y, Shareck M, Kestens Y. Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. *Prev Med* 2015; **71:** 83–7.

23 Mason KE, Bentley RJ, Kavanagh AM. Fruit and vegetable purchasing and the relative density of healthy and unhealthy food stores: evidence from an Australian multilevel study. *J Epidemiol Community Heal* 2013; **67:** 231–36.

24 Cannuscio CC, Tappe K, Hillier A, Buttenheim A, Karpyn A, Glanz K. Urban food environments and residents' shopping behaviors. *Am J Prev Med* 2013; **45:** 606–14.

25 Clary C, Lewis DJ, Flint E, Smith NR, Kestens Y, Cummins S. The local food environment and fruit and vegetable intake: a geographically weighted regression approach in the ORiEL Study. *Am J Epidemiol* 2016; **184:** 837–46.

26 Cummins S, Macintyre S. Food environments and obesity—neighbourhood or nation? *Int J Epidemiol* 2006; **35:** 100–04.

27 Mercille G, Richard L, Gauvin L, et al. The food environment and diet quality of urban-dwelling older women and men: Assessing the moderating role of diet knowledge. *Can J Public Heal* 2016; **107:** 34.

28 Stafford M, Cummins S, Macintyre S, Ellaway A, Marmot M. Gender differences in the associations between health and neighbourhood environment. *Soc Sci Med* 2005; **60:** 1681–92.

29 Simpson JA, MacInnis RJ, Peeters A, Hopper JL, Giles GG, English DR. A comparison of adiposity measures as predictors of all-cause mortality: the Melbourne Collaborative Cohort Study. *Obesity* 2007; **15:** 994–1003.

30 Fry A, Littlejohns T, Sudlow C, Doherty N, Allen N. OP41 The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. *J Epidemiol Community Health* 2016; **70:** A26.

4.3. Additional material

Neighbourhood greenspace and adiposity

Links between green space and health and wellbeing have long been recognised, but until recently rigorous scientific evidence on specific outcomes and the mechanisms involved has been limited. In the past decade there has emerged a substantial body of research investigating the relationship between routine exposure to green space, particularly in urban residential areas, and various health-related outcomes.

Various conceptual models linking green space to health have been proposed[1–3] and these typically recognise multiple pathways by which exposure to green space or nature more generally might influence health, including improved air quality, opportunities for physical activity, stress reduction and relaxation, resilience to heat-related illness by mitigation of the urban heat island effect, greater social cohesion, buffering from noise pollution, exposure to natural light, and improved functioning of the immune system[4].

Mostly because of the potential pathway through physical activity, green space is a potential determinant of overweight and obesity. Some of the other pathways identified above may also link green space to adiposity, for example through stress reduction or the positive effects of exposure to sunlight (e.g. improved sleep quality, which has been linked to metabolic outcomes including obesity[5]. The total influence of these pathways on obesity or adiposity can and has therefore been explored in a number studies[6], including one using UK Biobank[7] as well as many studies that have focussed on the relationship between greenspace and physical activity as either an outcome or a mediator of greenspace-health associations (e.g.[8–11]).

I have not focussed on the relationship between greenspace and adiposity in this thesis because other researchers using UK Biobank have already published results of a similar analysis using the greenspace measure available in UKBUMP (NDVI, which is only available for a subset of the cohort). The greenspace measure I use in Chapter 8 to examine associations with NCD outcomes did not become available until after the research paper in the current chapter was published. For completeness, I have repeated with the greenspace exposure the main, population-average analyses from the published paper, and I present those results here.

No association was observed between the amount of greenspace in the immediate neighbourhood and any of the adiposity outcomes, nor with the negative control, height (Table 4.5). I also checked whether the primary results of the published paper were sensitive to the addition of neighbourhood greenspace to the models (i.e. whether it may

have confounded the published analyses). Adding neighbourhood greenspace to the maximally adjusted models in the paper (for residents of England only due to greenspace data availability) made only negligible difference to the estimates for the formal physical activity environment. Estimates for proximity to a fast-food outlet were slightly attenuated, but upon further examination it was apparent that this was due to the restriction of the sample to residents of England, rather than adjustment for greenspace (not shown).

Table 4.5 Associations between neighbourhood greenspace and adiposity

	WC (cm)	BMI (kg/m2)	% body fat	Height (%) (negative control)
Greenspace + Gardens (300m buffer)	(n=355,471)	(n=355,049)	(n=349,890)	(n=355,249)
Least greenspace	ref	ref	ref	ref
Q2	0.35 (0.01, 0.69)	0.23 (0.10, 0.35)	0.25 (0.09, 0.42)	-0.11 (-0.21, -0.01)
Q3	0.22 (-0.20, 0.65)	0.21 (0.04, 0.39)	0.23 (0.04, 0.43)	-0.11 (-0.25, 0.02)
Q4	0.05 (-0.39, 0.49)	0.13 (-0.04, 0.29)	0.15 (-0.03, 0.33)	-0.04 (-0.19, 0.11)
Most greenspace	-0.27 (-0.76, 0.22)	0.01 (-0.17, 0.19)	-0.05 (-0.24, 0.14)	-0.01 (-0.16, 0.14)

ADDITIONAL REFERENCES

1 Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and Health. *Annu Rev Public Health* 2014; **35**: 207–28.

2 Villanueva K, Badland H, Hooper P, *et al.* Developing indicators of public open space to promote health and wellbeing in communities. *Appl Geogr* 2015; **57**: 112–9.

3 Lachowycz K, Jones AP. Towards a better understanding of the relationship between greenspace and health: development of a theoretical framework. *Landsc Urban Plan* 2013; **118**: 62–9.

4 World Health Organization. Urban green spaces and health: a review of the evidence. 2016.

5 Schmid SM, Hallschmid M, Schultes B. The metabolic burden of sleep loss. *Lancet Diabetes Endocrinol* 2015; **3**: 52–62.

6 Lachowycz K, Jones AP. Greenspace and obesity: a systematic review of the evidence. *Obes Rev* 2011; **12**: e183–9.

7 Sarkar C. Residential greenness and adiposity: findings from the UK Biobank. *Environ Int* 2017; **106**: 1–10.

8 Cummins S, Fagg J. Does greener mean thinner? Associations between neighbourhood greenspace and weight status among adults in England. *Int J Obes* 2012; **36**: 1108–13.

9 Richardson EA, Pearce J, Mitchell R, Kingham S. Role of physical activity in the relationship between urban green space and health. *Public Health* 2013; **127**: 318–24.

10 Villeneuve PJ, Jerrett M, Su JG, Weichenthal S, Sandler DP. Association of residential greenness with obesity and physical activity in a US cohort of women. *Environ Res* 2018; **160**: 372–84.

11 Schipperijn J, Cerin E, Adams MA, *et al.* Access to parks and physical activity: an eight country comparison. *Urban For Urban Green* 2017; **27**: 253–63.

Chapter 5. GENETIC RISK OF OBESITY AS A MODIFIER OF ASSOCIATIONS BETWEEN NEIGHBOURHOOD ENVIRONMENT AND BODY MASS INDEX

5.1. Introduction

In the previous chapter I presented evidence for characteristics of the physical activity and fast-food environments around people's homes being associated with adiposity, averaged across a study population of adults in mid-life, and also showed there were some gender- and income-related differences in the magnitude of those associations. These findings are consistent with some, but not all, other studies that have asked the same or similar research questions in other samples and other settings. In this next chapter I build upon those findings, extending – beyond gender and income differences – the idea that neighbourhood effects may not apply universally. Here in Chapter 5 I examine whether some individuals may be more sensitive to potentially obesogenic features of their local environment, depending on their underlying genetic risk of obesity. I present a research paper based on a novel study of gene-environment interactions, in which I have used the detailed genetic data in the UK Biobank resource to construct polygenic risk scores for obesity and tested whether these interact with measures of the neighbourhood food and PA environments to influence BMI. I also test interactions with a series of individual genetic variants linked to BMI through either diet or PA pathways, to check for consistency with the results of the polygenic risk score analyses.

Having improved the manuscript following a round of peer review with JAMA, I am now preparing the version included here for submission to *PLoS Medicine*.

5.2. Research Paper 2

Genetic risk of obesity as a modifier of associations between neighbourhood environment and Body Mass Index

Note: Supplementary material for this research paper is included in Appendix Two.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1510923	**Title**	Ms
First Name(s)	Kate		
Surname/Family Name	Mason		
Thesis Title	Where and for whom does the neighbourhood built environment matter for obesity and health?: Examining sources of effect heterogeneity at multiple scales in the UK adult population		
Primary Supervisor	Neil Pearce		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	PLoS Medicine
Please list the paper's authors in the intended authorship order:	Kate Mason, Luigi Palla, Neil Pearce, Jody Phelan, Steven Cummins
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the analysis with input from LP and SC, and independently undertook the data management, statistical analysis, and writing of the manuscript. JP extracted the genetic data from UK Biobank and JP and LP provided expertise on working with genetic data. SC, LP and NP contributed to the interpretation of results, and final drafting of the manuscript.

SECTION E

Student Signature	[redacted]
Date	01/07/2019

Supervisor Signature	[redacted]
Date	01/07/2019

Genetic risk of obesity as a modifier of associations between neighbourhood environment and Body Mass Index

ABSTRACT

Background: There is growing recognition that recent global increases in obesity are the product of a complex interplay between genetic and environmental factors. However, in gene-environment studies of obesity, 'environment' usually refers to individual behavioural factors that influence energy balance, while more upstream environmental factors are overlooked. We examined gene-environment interactions between genetic risk of obesity and two neighbourhood characteristics likely to be associated with obesity (proximity to takeaway/fast-food outlets and availability of physical activity facilities) and genetic risk of obesity.

Methods: We used data from 335,046 adults aged 40-70 in the UK Biobank cohort to conduct a population-based cross-sectional study of interactions between neighbourhood characteristics and genetic risk of obesity, in relation to BMI. Proximity to a fast-food outlet was defined as distance from home address to nearest takeaway/fast-food outlet, and availability of physical activity facilities as the number of formal physical activity facilities within one kilometre of home address. Genetic risk of obesity was operationalised by 91-SNP and 69-SNP weighted genetic risk scores, and by six individual SNPs considered separately. Multivariable, mixed effects models with product terms for the gene-environment interactions were estimated.

Results: After accounting for likely confounding, the association between proximity to takeaway/fast-food outlets and BMI was stronger among those at increased genetic risk of obesity, with evidence of an interaction with polygenic risk scores (P=0.018 and P=0.028 for 69-SNP and 91-SNP scores, respectively) and in particular with a SNP linked to *MC4R* (P=0.009), a gene known to regulate food intake. We found no evidence of a gene-environment interaction for availability of physical activity facilities.

Conclusions: Individuals at an increased genetic risk of obesity may be more sensitive to exposure to the local fast-food environment. Ensuring that neighbourhood residential environments are designed to promote a healthy weight may be particularly important for those with greater genetic susceptibility to obesity.

BACKGROUND

Obesity has a heritable component[1], but the rapid rise in global obesity prevalence suggests an important role for environmental influences[2]. However, individuals may have differing physiological or behavioural responses to the increasingly 'obesogenic' environment, suggesting that a complex interplay between genetic and non-genetic factors affects weight[3,4].

Advances in genotyping technologies have enabled the investigation of gene-environment (GxE) interactions[4,5]. For obesity outcomes, the 'environment' in GxE studies is often operationalised as the lifestyle or behavioural factors that influence energy balance[6], rather than more upstream features of the built and social environments; the settings where behavioural 'choices' are made and constrained. Despite long being recognised in social epidemiology as potentially important determinants of weight status, these 'socio-ecological' environmental factors have been examined in only a small number of GxE studies[7–12].

The residential neighbourhood environment comprises many features that potentially influence energy balance. These include the proximity, density and relative proportions of healthy and unhealthy food retailers[13–15], and resources for physical activity (PA), such as leisure centres, swimming pools, gyms and sports fields[16–19]. Other neighbourhood features linked to energy balance include walkability, access to public transport and local resources such as public parks and greenspace[20,21]. If the genetic risk of obesity modifies the influence of these neighbourhood exposures, we would expect to observe differential effects of the residential environment on body mass index (BMI) according to level of genetic risk. The influence of the environment may be strongest in people with high genetic risk due to increased sensitivity to external factors[22,23], or it may be strongest in people with low genetic risk, who maximise their genetic 'advantage' within a healthier environment while those at greater risk express a higher BMI phenotype regardless of environmental factors[6].

In this study we use the UK Biobank cohort to examine whether genetic risk of obesity modifies the effect of two residential environment exposures likely to influence BMI: proximity to fast-food and availability of formal PA facilities. We operationalise genetic risk in two ways. First, using polygenic risk scores derived from single nucleotide polymorphisms (SNPs) linked to BMI, and second, using the individual SNPs most strongly linked to BMI and thought to be involved in diet or PA pathways.

METHODS

Data

We used baseline data from UK Biobank[24]. Data were potentially available from 502,656 individuals who visited 22 UK Biobank assessment centres across the UK between 2006 and 2010. Individuals aged 40–69 years living within 25 miles of an assessment centre and listed on National Health Service (NHS) patient registers were invited to participate.

Linked to UK Biobank is the UK Biobank Urban Morphometric Platform (UKBUMP), a high-resolution spatial database of objectively measured characteristics of the physical environment surrounding each participant's residential address, derived from multiple national spatial datasets[25]. Environmental measures include densities of various land uses and proximity to various health-relevant resources. Measures for the current study are available for 96% of the UK Biobank sample.

Genome-wide genetic data are available for 488,363 participants. Genetic data are missing from 3% of the sample as insufficient DNA was extracted from blood samples for genotyping assays. SNP genotypes not directly assayed were imputed. Procedures used to derive the genetic data and undertake quality assurance are reported in Bycroft et al[26]. Genetic data for the relevant SNPs were downloaded, decrypted and linked to participant IDs to facilitate analysis.

Outcome

Body Mass Index (BMI, kg/m^2) was calculated from weight and height measurements collected by trained staff using standard procedures[24]. The variable was normally distributed and analysed as a continuous outcome variable.

Neighbourhood exposures

We examined interactions between genetic risk and two neighbourhood characteristics likely to influence BMI: availability of formal PA facilities (number of indoor and outdoor sporting and leisure facilities within a one-kilometre street-network distance of an individual's home) and fast-food proximity (distance in metres to nearest takeaway/fast-food outlet). Greater neighbourhood availability of PA facilities may influence BMI through increased opportunities for physical activity, and greater distances from home to fast-food outlets may influence BMI by reducing access to fast food[27,28]. In prior analyses we found both were associated with BMI in the expected direction – that is, living further

from a fast-food outlet, or having more PA facilities near home, was associated with having a lower BMI[16]. Both exposures were analysed as continuous variables, with higher values of each (more facilities; greater distance to nearest fast-food outlet) representing lower exposure. Due to the positively skewed distribution of these variables, number of PA facilities was capped at 15 (<1% recoded from >15) and distance to nearest fast-food outlet was log transformed (base 10) such that regression coefficients were interpreted as the mean difference in BMI associated with a 10-fold increase in distance to nearest fast-food outlet e.g. 100 metres to one kilometre.

Genetic risk scores and individual SNPs

A recent genome-wide association study (GWAS) identified 97 SNPs associated with BMI[29]. We constructed a genetic risk score (GRS) based on 91 of these SNPS, excluding six SNPs identified elsewhere[30] as being in linkage disequilibrium with other included SNPs (rs17001654, rs2075650 and rs9925964) or having pleiotropic effects (rs11030104, rs3888190, rs13107325), both of which may produce bias in associations between the genetic risk score and the outcome, and in interaction analyses[31]. We also constructed an alternative GRS, the same as one by used Tyrrell and colleagues[30] in a study of UK Biobank participants of White British ancestry, in which they tested interactions between genetic risk and behavioural exposures using a GRS derived from 69 of the SNPs identified in the recent GWAS. Their GRS excluded SNPs from secondary meta-analyses of studies of regional, sex-stratified or non-European-descent populations[29], and one SNP (rs2033529) that was unavailable at the time of their study. Full lists of the SNPs included in each of the 91-SNP and 69-SNP risk scores are provided in Supplementary Table 4 (Appendix Two). The GRSs were constructed by summing the number of BMI-increasing alleles across the set of 69 or 91 loci, and weighting the allele count at each SNP by its published effect size[29]. For imputed SNP genotypes we used the imputed allelic dosages.

From the literature we identified individual SNPs with a well-established link to obesity and the largest published effect sizes (rs1558902 rs6567160 rs13021737, markers of the *FTO*, *MC4R* and *TMEM18* genes respectively)[1,29], and three SNPs recently linked to physical activity (rs13078960, rs10938397 and rs7141420, markers of *CADM2*, *GNPDA2 NRXN3*)[32,33]. We tested for interactions between the number of BMI-increasing alleles at each of these loci, and each neighbourhood variable. We hypothesise that if GxE interactions are observed for these SNPs, those SNPs implicated in dietary behaviour will only interact with the fast-food environment, and those implicated in PA behaviour will only interact with the PA environment.

Covariates

Models were adjusted for potential confounding by age, sex, educational attainment, household income, employment status, area deprivation (Townsend score), urban/non-urban status, and neighbourhood residential density and mutually adjusted for the other neighbourhood exposure. We also corrected for population stratification by adjusting for the first ten of 40 UK Biobank-provided genetic ancestry principal components from a genome-wide PCA of UK Biobank's genetic data[26].

Statistical analysis & analytic sample

Accounting for the nested structure of the data (individuals within assessment areas), we used mixed effects models with a random coefficient for the neighbourhood exposure and assuming an unstructured variance/covariance matrix. Models included an interaction term between the neighbourhood exposure and the genetic risk score, with both analysed as continuous variables. BMI difference per unit change in the exposure was estimated for each quintile of genetic risk. The *p*-value for the additive interaction term was interpreted as strength of evidence of effect modification. The marginal predicted values of BMI associated with different levels of each neighbourhood exposure from these models were plotted for the top and bottom quintile of genetic risk, to visualise observed effect heterogeneity according to genetic risk. A complete case analysis was used, restricted to UK Biobank participants of White British ancestry (defined by concordant self-report and PCA results for White British/Caucasian ancestry) for the primary analyses (N=335,046) because the smaller GRS was limited to SNPs associated with BMI in analyses of individuals with European ancestry. Analysis was performed using Stata SE v14.2 (Stata Corp, Texas USA).

Sensitivity analyses

As the 91-SNP GRS included SNPs associated with BMI in populations of non-European descent, we undertook a sensitivity analysis that tested for an interaction with the 91-SNP GRS in a sample unrestricted by ethnicity to test generalisability to the wider source population. To explore the possibility that results might be biased by latent genetic structure in the sample – a concern regarding genetic analyses involving UK Biobank[34] – we also performed sensitivity analyses in which models were adjusted for all 40 genetic ancestry principal components, and for birth location. Finally, although weighting of the polygenic risk scores is appropriate due to the varying degree to which each SNP is associated with BMI, we performed sensitivity analyses using an unweighted version of

each GRS. Evidence of a GxE interaction using unweighted scores is expected to be weaker, due to dilution of the effects of the more influential SNPs.

Ethics

UK Biobank has ethical approval from the North West Multi-centre Research Ethics Committee (reference 16/NW/0274), the Patient Information Advisory Group (PIAG), and the Community Health Index Advisory Group (CHIAG). Additional ethical approval for the specific study was obtained from the London School of Hygiene and Tropical Medicine's Research Ethics Committee in September 2016 (reference 11897).

RESULTS

The sample was 52.2% female, with a mean age of 56.5 years (range 40-70 years at baseline). Mean BMI was 27.4 kg/m² (SD=4.7), median distance to nearest fast-food outlet was 1171 metres and median number of PA facilities within one kilometre of home was one. Sample characteristics are summarised in Table 5.1.

Using the two alternative weighted genetic risk scores, we observed evidence of an interaction between fast-food proximity and genetic risk (P=0.028 for the 91-SNP GRS, P=0.018 for the 69-SNP GRS). The magnitude of the estimated effect between fast-food proximity and BMI was small at all levels of genetic risk, but increased as genetic risk increased. In the highest quintile of genetic risk of obesity (based on the 91-SNP GRS), each 10-fold increase in distance to the nearest fast-food store was associated with a 0.194kg/m² lower mean BMI (95%CI: -0.326,-0.062), which was twice the magnitude of association in the lowest risk quintile (β=-0.081; 95%CI: -0.213,0.052) (Table 5.2; Figure 5.1).

Table 5.1 Characteristics of total sample and top and bottom quintile of 91-SNP genetic risk score

	91-SNP genetic risk score		Total sample
	Quintile 1 (lowest risk of obesity)	Quintile 5 (highest risk of obesity)	
Total number of participants	64269	69577	335046
BMI (kg/m^2) *(mean, SD)*	26.5 (4.3)	28.3 (5.1)	27.4 (4.7)
Distance to nearest fast-food outlet (m) *(median, IQR)*	1172 (634 - 2302)	1169 (626 - 2290)	1171 (630-2301)
Number of PA facilities within 1km of home address *(median, IQR)*	1 (0 - 3)	1 (0 - 3)	1 (0 - 3)
Age *(mean, SD)*	56.5 (8.0)	56.5 (8.0)	56.5 (8.0)
Sex (female)	33876 (52.7%)	35923 (51.6%)	174872 (52.2%)
Income			
Less than 18,000	14154 (22.0%)	15734 (22.6%)	74556 (22.3%)
18,000 to 30,999	16497 (25.7%)	18270 (26.3%)	86917 (25.9%)
31,000 to 51,999	17013 (26.5%)	18374 (26.4%)	88721 (26.5%)
52,000 to 100,000	13269 (20.7%)	13738 (19.8%)	67908 (20.3%)
Greater than 100,000	3336 (5.2%)	3461 (5.0%)	16944 (5.1%)
Education			
College or University degree	21462 (33.4%)	22412 (32.2%)	110153 (32.9%)
A levels/AS levels or equivalent	7635 (11.9%)	7961 (11.4%)	39017 (11.7%)
O levels/GCSEs or equivalent	14262 (22.2%)	15779 (22.7%)	74966 (22.4%)
CSEs or equivalent	3500 (5.5%)	3933 (5.7%)	18722 (5.6%)
NVQ or HND or HNC or equivalent	4266 (6.6%)	4782 (6.9%)	22892 (6.8%)
Other professional qualifications	3302 (5.1%)	3527 (5.1%)	16954 (5.1%)
None of the above	9842 (15.3%)	11183 (16.1%)	52342 (15.6%)
Employment status			
Paid employment or self-employed	38217 (59.5%)	41326 (59.4%)	199280 (59.5%)
Retired	21330 (33.2%)	23096 (33.2%)	111113 (33.2%)
Unable to work	1756 (2.7%)	2055 (3.0%)	9457 (2.8%)
Unemployed	769 (1.2%)	878 (1.3%)	4238 (1.3%)
Home duties/carer/student/other	2197 (3.4%)	2222 (3.2%)	10958 (3.3%)
Urbanicity (% urban dwelling)	54560 (84.9%)	59018 (84.8%)	284471 (84.9%)
Area deprivation[a] *(mean, SD)*	-1.6 (2.9)	-1.6 (2.9)	-1.6 (2.9)
Residential density[b] *(median, IQR)*	1794 (1041 - 2934)	1801 (1043 - 2911)	1798 (1044 - 2918)

[a] 2001 Townsend index score

[b] Residential address points per 1km street-network buffer

There was less evidence that the association between availability of PA facilities and BMI was modified by genetic risk. The magnitude of the association between number of formal PA facilities within 1km of home and BMI was similar at all levels of genetic risk, and while effect estimates did increase slightly with increasing genetic risk, differences between risk groups were small with little evidence of interaction for both the 91-SNP GRS (P=0.530) and 69-SNP GRS (P=0.178). For both environmental exposures, the results obtained from the two different weighted GRSs were substantively similar, but evidence of an interaction with the 69-SNP GRS was somewhat stronger (Table 5.2;Figure 5.1). The plots in Figure 5.1 also demonstrate that the BMI difference between the highest and lowest risk quintiles is greater for the 91-SNP GRS than the 69-SNP GRS, reflecting the fact that the larger GRSs captures more of the genetic variation in BMI.

Examination of interactions between neighbourhood variables and specific SNPs revealed strong evidence of one interaction: among people with higher risk allele counts at the marker of *MC4R*, which encodes the melanocortin-4 receptor previously shown to be important in the regulation of food intake, living nearer to a fast-food store was more closely associated with higher BMIs than it was among people with fewer risk alleles at this locus ($P_{interaction}$=0.009; Table 5.3 and Figure 5.2). Some evidence of an interaction between fast-food proximity and rs1558902, the marker of the *FTO* gene (P=0.067), where again the higher risk group showed a stronger association between fast-food proximity and BMI was observed. We also observed some evidence of a GxE interaction between the availability of PA facilities and rs13021737 (in the *TMEM18* gene) (P=0.076). In this case, increased genetic risk attenuated the association between availability of PA facilities and BMI slightly, but the difference in magnitude between high and low risk groups was small (Figure 5.2).

Table 5.2 Associations between neighbourhood variables and BMI, by quintile of genetic risk based on 91-SNP and 69-SNP risk scores

	91-SNP GRS			69-SNP GRS		
	Quintile of genetic risk	Mean BMI difference for unit increase in neighbourhood exposure	P-interaction	Quintile of genetic risk	Mean BMI difference for unit increase in neighbourhood exposure	P-interaction
Fast-food proximity[a,b] (beta represents BMI difference for a 10-fold increase in distance (m) to nearest fast-food outlet)	Q1	-0.081 (-0.213, 0.052)	0.028	Q1	-0.080 (-0.214, 0.055)	0.018
	Q2	-0.115 (-0.239, 0.009)		Q2	-0.117 (-0.243, 0.009)	
	Q3	-0.137 (-0.259, -0.014)		Q3	-0.140 (-0.264, -0.017)	
	Q4	-0.158 (-0.282, -0.035)		Q4	-0.164 (-0.289, -0.039)	
	Q5	-0.194 (-0.326, -0.062)		Q5	-0.204 (-0.337, -0.070)	
Availability of PA facilities[a,c] (beta represents BMI difference for each additional facility)	Q1	-0.074 (-0.100, -0.047)	0.530	Q1	-0.070 (-0.097, -0.044)	0.178
	Q2	-0.075 (-0.101, -0.049)		Q2	-0.074 (-0.099, -0.048)	
	Q3	-0.076 (-0.101, -0.050)		Q3	-0.076 (-0.101, -0.050)	
	Q4	-0.077 (-0.103, -0.051)		Q4	-0.078 (-0.103, -0.052)	
	Q5	-0.078 (-0.105, -0.052)		Q5	-0.081 (-0.106, -0.054)	

[a] Regression models were adjusted for age (years), sex (male/female), highest education level attained (Degree; A level or equivalent; O level or equivalent; CSE or equivalent; NVQ/HND/HNC; other professional qualification; none of the above), annual household income (<£18,000; £18,000-£30,999; £31,000-£51,999; £52,000-£100,000; >£100,000), employment status (paid work, retired, unable to work, unemployed, other), area deprivation (Townsend score), urbanicity (urban/non-urban), neighbourhood residential density (count of residential features within a one-km street network buffer of home address, log transformed).

[b] Also adjusted for availability of PA facilities.

[c] Also adjusted for fast-food proximity.

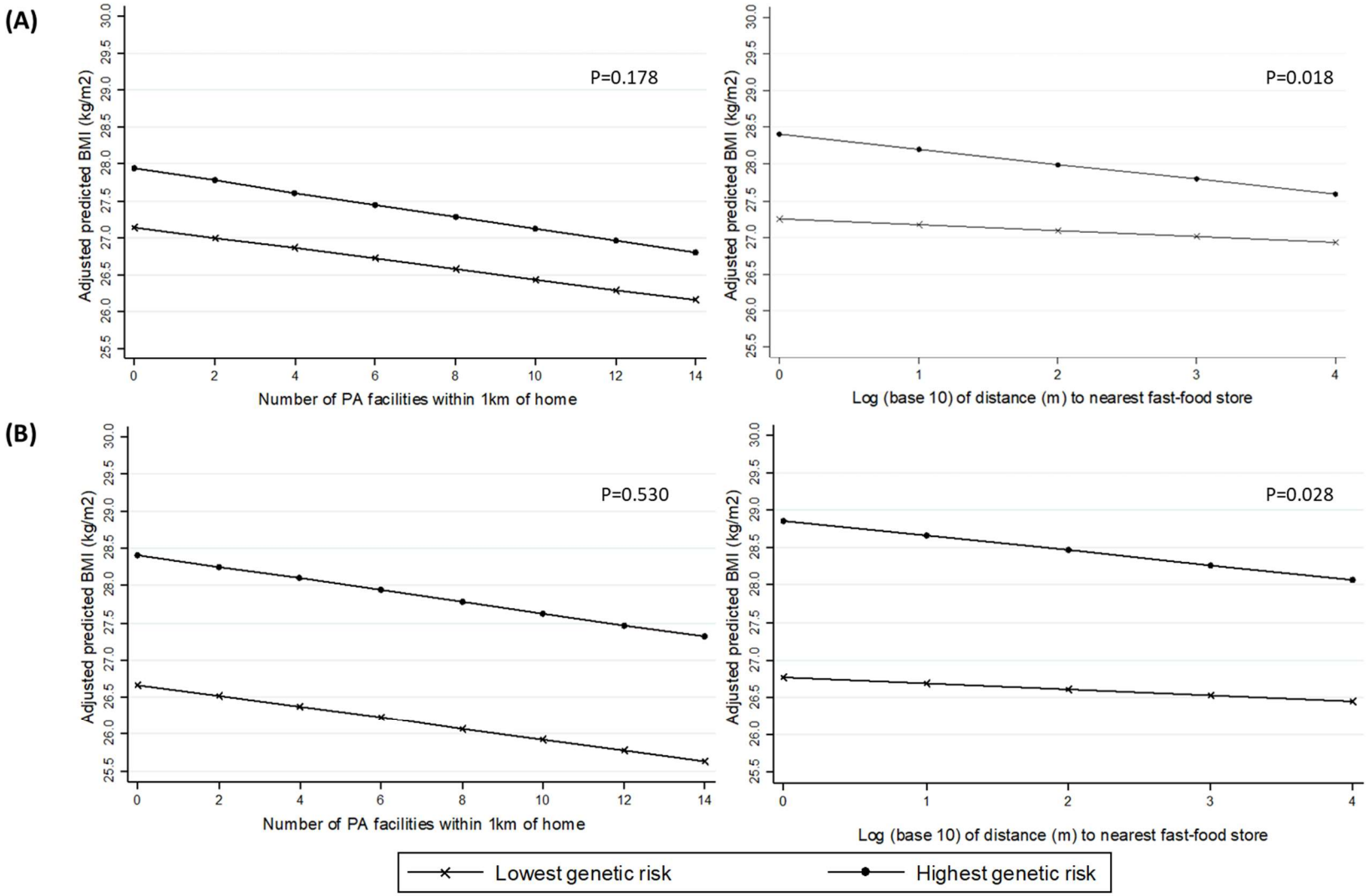


Figure 5.1 Association between neighbourhood variables and BMI in the highest and lowest quintiles of genetic risk, based on (a) 69-SNP Genetic Risk Score, and (b) 91-SNP Genetic Risk Score

Table 5.3 Association between neighbourhood variables and BMI, testing interaction with number of risk alleles at selected loci

		rs1558902 *(FTO)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.067	-0.099 (-0.198, -0.001)	-0.148 (-0.238, -0.059)	-0.197 (-0.305, -0.088)
PA facilities	0.933	-0.077 (-0.104, -0.050)	-0.077 (-0.103, -0.051)	-0.076 (-0.104, -0.049)
		rs6567160 *(MC4R)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.009	-0.096 (-0.188, -0.003)	-0.177 (-0.271, -0.083)	-0.258 (-0.386, -0.130)
PA facilities	0.606	-0.078 (-0.104, -0.051)	-0.075 (-0.102, -0.049)	-0.073 (-0.103, -0.043)
		rs13021737 *(TMEM18)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.993	-0.135 (-0.226, -0.043)	-0.135 (-0.234, -0.036)	-0.135 (-0.279, 0.008)
PA facilities	0.076	-0.080 (-0.106, -0.053)	-0.071 (-0.098, -0.043)	-0.061 (-0.093, -0.030)
		rs13078960 *(CADM2)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.114	-0.159 (-0.252, -0.066)	-0.108 (-0.205, -0.010)	-0.056 (-0.192, 0.081)
PA facilities	0.419	-0.076 (-0.102, -0.049)	-0.079 (-0.106, -0.053)	-0.083 (-0.114, -0.053)
		rs10938397 *(GNPDA2)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.328	-0.115 (-0.215, -0.015)	-0.141 (-0.230, -0.052)	-0.167 (-0.274, -0.061)
PA facilities	0.694	-0.076 (-0.102, -0.049)	-0.077 (-0.103, -0.051)	-0.079 (-0.106, -0.051)
		rs7141420 *(NRXN3)*		
	P-interaction	Homozygous low risk (0 risk alleles)	Heterozygous (1 risk allele)	Homozygous high risk (2 risk alleles)
Fast-food proximity	0.520	-0.152 (-0.257, -0.048)	-0.135 (-0.227, -0.043)	-0.118 (-0.224, -0.012)
PA facilities	0.125	-0.071 (-0.097, -0.044)	-0.077 (-0.102, -0.051)	-0.083 (-0.110, -0.056)

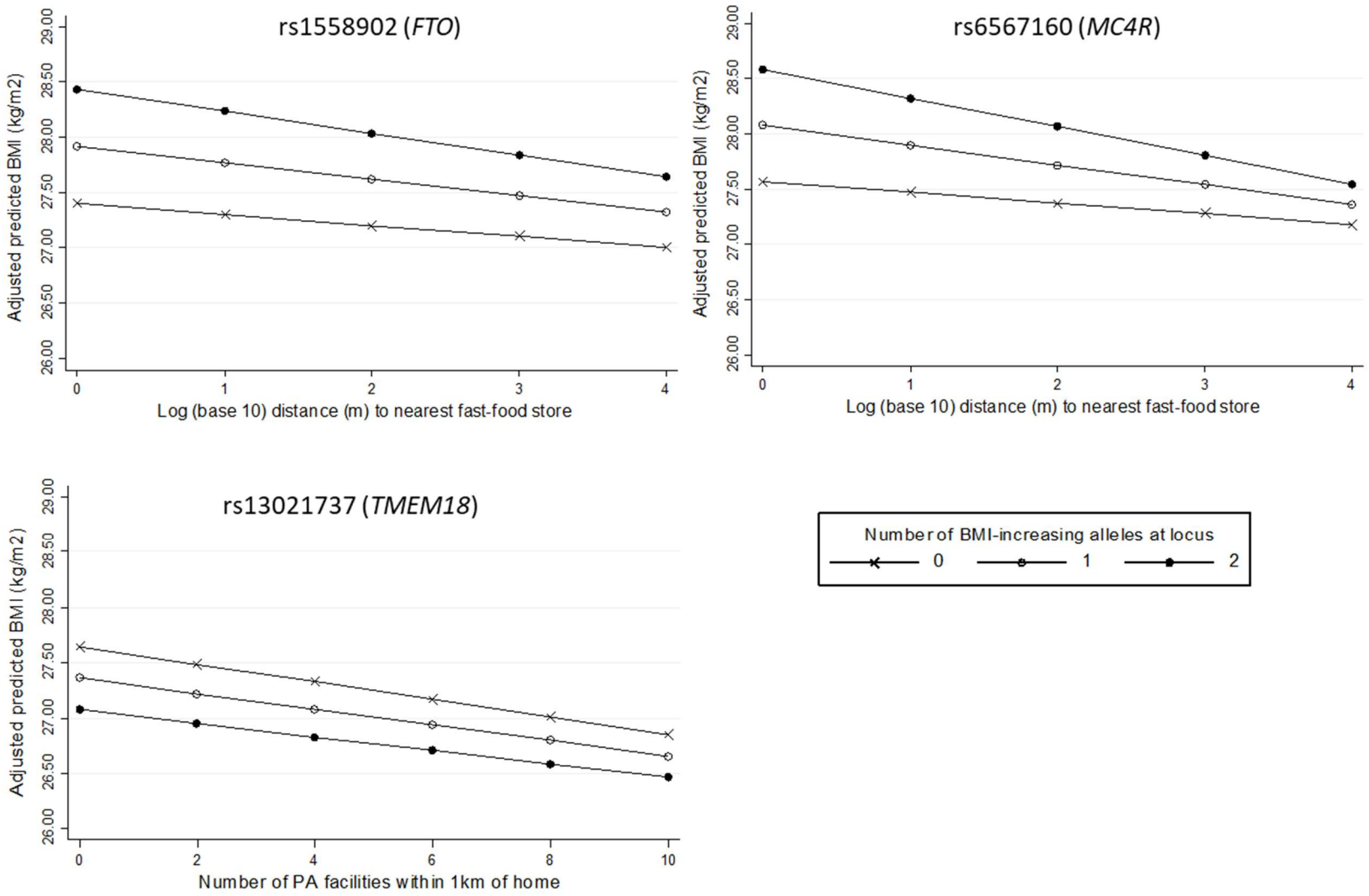


Figure 5.2 Association between neighbourhood variables and BMI according to number of risk alleles at individual SNPs where $P_{interaction}$<0.10 (rs1558902 & rs6567160 for fast-food proximity; rs13021737 for availability of PA facilities)

In sensitivity analyses, interactions between fast-food proximity and genetic risk were – as expected – weaker when the genetic risk scores were not weighted by the effect sizes of the component SNPs, with mean differences in BMI more similar across levels of genetic risk than we observed using the weighted score (Supplementary Table 5). Expanding the sample to include non-White ethnicities, we observed slightly increased P-values for the interaction terms but otherwise no substantive difference from the primary analysis (Supplementary Table 6). For all models, the impact of adjusting for 40 rather than 10 genetic ancestry principal components was negligible, while some attenuation of the interaction between fast-food proximity and polygenic risk occurred when adjusting for birth location (Supplementary Table 7).

DISCUSSION

In UK Biobank we found evidence that genetic risk of obesity modifies sensitivity to the neighbourhood food environment, though effects are small. We found that people at higher genetic risk of obesity have higher average BMI the closer they live to a fast-food outlet, whereas for those at low genetic risk of obesity, distance to the nearest fast-food outlet does not appear to be associated with BMI. In contrast, an overall negative association between neighbourhood availability of PA facilities and BMI varies very little across levels of polygenic risk.

The observed gene-environment interaction for fast-food proximity using polygenic risk scores was supported by stronger evidence of an interaction between fast-food proximity and a specific SNP near *MC4R*, a gene known to be involved in regulation of food intake[35]. Previous research has linked *MC4R* specifically to binge eating[36] although this remains contested[37]. We also observed some evidence of a possible interaction with a SNP marker of *FTO*, a gene with well-established links to obesity. While *FTO* has long been recognised as an obesity-associated locus, and has been implicated in central nervous system regulation of appetite, its exact function remains poorly understood[1]. In a study of gene-diet interactions, genetic risk scores for BMI were found to be associated with fried food consumption, and, consistent with our results, individual loci in or near both *MC4R* and *FTO* contributed to this[38].

Limited evidence for an interaction between genetic risk and the PA environment is consistent with findings from a recent study in adolescents that found that availability of recreation facilities did not contribute to the attenuation by PA of genetic risk of obesity[32].

While overall genetic risk of obesity did not interact with the PA environment in our study, the weaker association we observed between the availability of PA facilities and BMI in those with more risk alleles at the *TMEM18* locus suggests that some specific SNPs might. Further examination of other SNPs is warranted. Lack of interaction with specific SNPs might be explained by the pathways they influence being less sensitive to environmental exposures. As the functional pathways by which most BMI-associated loci influence BMI remain poorly understood, it is difficult to speculate further.

Stronger evidence for interactions with specific SNPs highlights the lack of specificity of polygenic risk scores. While useful in exploratory studies, grouping all SNPs statistically associated with a complex phenotype such as BMI into a single score, regardless of the function of the genes they represent, may dilute or obscure important interactions. Scores based on known or putative biological mechanisms may prove more valuable, particularly for elucidating causal relationships. We observed very similar results for both the 69-SNP and 91-SNP genetic risk scores, although the smaller GRS yielded stronger evidence of interaction. It may be that the additional SNPs in the larger GRS diluted the interaction due to being associated with BMI only in some population subgroups, and some having been linked to BMI only in more ethnically diverse populations than our primary sample.

We have reported elsewhere that the main association between fast-food proximity and BMI in UK Biobank may be attenuated due to measurement error in the exposure[16], and because the exposure does not account for other, healthier elements of the food environment[39]. Compared with other measures of the fast-food environment, proximity measures may also produce more conservative estimates of association with relevant outcomes[40]. In a regional sub-sample of UK Biobank, others have recently improved on this measurement of the food environment and found stronger associations[41]. In this study, where the main effect sizes are relatively small, even the reasonably strong interaction effects we observed translate to small differences between high and low risk groups. However, given the likely measurement error and the distal and complex nature of the relationships under investigation, detecting even weak associations and small differences might point to potentially important processes. Here we examined only two characteristics of neighbourhood environments; others may also interact with genetic risk. For example, GxE interactions have recently been reported for neighbourhood walkability and obesity[11], and neighbourhood deprivation and BMI[10]. Given that unhealthy characteristics of neighbourhoods often cluster together[42], the combined effects of

multiple 'obesogenic' features on those at increased genetic risk of obesity may be substantial.

Our findings provide evidence for a potentially important GxE interaction, but further confirmatory studies are required. Another recent study found a strong GxE interaction between genetic risk of obesity and socioeconomic status, and while our analyses are adjusted for several socioeconomic indicators, if there remains any residual confounding by SES it may be contributing to the GxE interactions we observed. Geographical genetic structure in the sample remains a risk, even after adjustment for ancestry components and geography. Such structure may induce spurious associations with polygenic risk scores in particular[34]. In sensitivity analyses we found that adjustment for additional ancestry principal components had negligible impact on the strength of evidence for the GxE interactions we tested, but evidence for a genetic interaction with fast-food proximity was slightly weaker following adjustment for birth location. Further investigation of the effect of the residual genetic structure in the sample is warranted. GxE interactions are also sensitive to the scaling of environmental variables, and the power to detect a GxE interaction can depend on the main effect sizes, and distribution and measurement quality of the genetic and environmental variables[43]. Studies using UK Biobank are also at risk of selection bias due to a low response rate[44]. It is important these analyses are replicated in other samples at lower risk of these biases.

It is widely accepted that environmental factors are important in explaining the recent rise in the global prevalence of overweight and obesity. In this study, we find evidence that people at higher genetic risk of obesity may more sensitive to exposure to the residential fast-food environment. Ensuring that neighbourhood residential environments are designed to promote a healthy weight may be particularly important for those with genetic susceptibility to obesity.

REFERENCES

1 Waalen J. The genetics of human obesity. *Transl Res* 2014; **164**: 293–301.

2 Popkin BM, Gordon-Larsen P. The nutrition transition: worldwide obesity dynamics and their determinants. *Int J Obes Relat Metab Disord* 2004; **28 Suppl 3**: S2-9.

3 Van Der Klaauw AA, Farooqi IS. The hunger genes: pathways to obesity. *Cell* 2015; **161**: 119–32.

4 Richmond RC, Timpson NJ. Recent findings on the genetics of obesity: is there public health relevance? *Curr Nutr Rep* 2012; **1**: 239–48.

5 Andreasen CH, Andersen G. Gene-environment interactions and obesity: further aspects of genomewide association studies. *Nutrition* 2009; **25**: 998–1003.

6 Boardman JD, Daw J, Freese J. Defining the environment in gene-environment research: lessons from social epidemiology. *Am J Public Health* 2013; **103**: S64-72.

7 Horn EE, Turkheimer E, Strachan E, Duncan GE. Behavioral and environmental modification of the genetic influence on body mass index: a twin study. *Behav Genet* 2015; **45**: 409–26.

8 Liu H, Guo G. Lifetime socioeconomic status, historical context, and genetic inheritance in shaping body mass in middle and late adulthood. *Am Sociol Rev* 2015; **80**: 705–37.

9 Mooney SJ, Grady ST, Sotoodehnia N, *et al.* In the wrong place with the wrong SNP: the association between stressful neighborhoods and cardiac arrest within beta-2-adrenergic receptor variants. *Epidemiology* 2016; **27**: 656-62

10 Owen G, Jones K, Harris R. Does neighbourhood deprivation affect the genetic influence on body mass? *Soc Sci Med* 2017; **185**: 38–45.

11 Kowaleski-Jones L, Brown BB, Fan JX, Hanson HA, Smith KR, Zick CD. The joint effects of family risk of obesity and neighborhood environment on obesity among women. *Soc Sci Med* 2017; **195**: 17–24.

12 Robinette JW, Boardman JD, Crimmins EM. Differential vulnerability to neighbourhood disorder: a gene×environment interaction study. *J Epidemiol Community Health* 2019; : jech-2018-211373.

13 Fleischhacker SE, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. *Obes Rev* 2011; **12**: e460-71.

14 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 2014; **348**: g1464.

15 Black C, Moon G, Baird J. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health Place* 2014; **27**: 229–42.

16 Mason KE, Pearce N, Cummins S. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *Lancet Public Heal* 2018; **3**: e24–33.

17 Ellaway A, Lamb KE, Ferguson NS, Ogilvie D. Associations between access to recreational physical activity facilities and body mass index in Scottish adults. *BMC Public Health* 2016; **16**: 756.

18 Van Holle V, Deforche B, Van Cauwenberg J, *et al.* Relationship between the physical environment and different domains of physical activity in European adults: a systematic review. *BMC Public Health* 2012; **12**: 807.

19 Barrientos-Gutierrez T, Moore KAB, Auchincloss AH, *et al.* Neighborhood physical environment and changes in body mass index: results from the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 2017; **186**: 1–9.

20 Townshend T, Lake AA. Obesogenic urban form: theory, policy and practice. *Health Place* 2009; **15**: 909–16.

21 Black JL, Macinko J. Neighborhoods and obesity. *Nutr Rev* 2008; **66**: 2–20.

22 Davis OSP, Haworth CMA, Lewis CM, Plomin R. Visual analysis of geocoded twin data puts nature and nurture on the map. *Mol Psychiatry* 2012; **17**: 867–74.

23 Llewellyn C, Wardle J. Behavioral susceptibility to obesity: gene-environment interplay in the development of weight. *Physiol Behav* 2015; **152**: 494–501.

24 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007.

25 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP): a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21**: 135–48.

26 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**: 203–9.

27 Block J, Seward M, James P. Food environment and health. In: Duncan DT, Kawachi I, eds. Neighborhoods and Health, 2nd edn. New York: Oxford University Press, 2018: 247–77.

28 Sallis JF, Glanz K. Physical activity and food environments: solutions to the obesity epidemic. *Milbank Q* 2009; **87**: 123–54.

29 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.

30 Tyrrell J, Wood AR, Ames RM, *et al.* Gene-obesogenic environment interactions in the UK Biobank study. *Int J Epidemiol* 2017; **46**: 559–75.

31 Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet* 2014; **95**: 301–7.

32 Graff M, Richardson AS, Young KL, *et al.* The interaction between physical activity and obesity gene variants in association with BMI: does the obesogenic environment matter? *Health Place* 2016; **42**: 159–65.

33 Klimentidis YC, Raichlen DA, Bea J, *et al.* Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including *CADM2* and *APOE*. *Int J Obes* 2018; **42**: 1161–76.

34 Haworth S, Mitchell R, Corbin L, *et al.* Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019; **10**: 333.

35 Spiegelman BM, Flier JS. Obesity and the regulation of energy balance. *Cell* 2001; **104**: 531–43.

36 Branson R, Potoczna N, Kral JG, Lentes K-U, Hoehe MR, Horber FF. Binge eating as a major phenotype of melanocortin 4 receptor gene mutations. *N Engl J Med* 2003; **348**: 1096–103.

37 Valette M, Bellisle F, Carette C, *et al.* Eating behaviour in obese patients with melanocortin-4 receptor mutations: a literature review. *Int J Obes* 2013; **37**: 1027–35.

38 Qi Q, Chu AY, Kang JH, *et al.* Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ* 2014; **348**: g1610.

39 Cummins S, Clary C, Shareck M. Enduring challenges in estimating the effect of the food environment on obesity. *Am J Clin Nutr* 2017; **106**: 445–6.

40 Bivoltsis A, Cervigni E, Trapp G, Knuiman M, Hooper P, Ambrosini GL. Food environments and dietary intakes among adults: does the type of spatial exposure measurement matter? A systematic review. *Int J Health Geogr* 2018; **17**: 1–20.

41 Burgoine T, Sarkar C, Webster CJ, Monsivais P. Examining the interaction of fast-food outlet exposure and income on diet and obesity: evidence from 51,361 UK Biobank participants. *Int J Behav Nutr Phys Act* 2018; **15**: 1–12.

42 Fraser LK, Edwards KL, Cade J, Clarke GP. The geography of fast food outlets: a review. *Int J Environ Res Public Health* 2010; **7**: 2290–308.

43 Manuck SB, McCaffery JM. Gene-environment interaction. *Annu Rev Psychol* 2014; **65**: 41–70.

44 Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2017; : 1–10.

Chapter 6. INTERACTIONS BETWEEN NEIGHBOURHOOD CHARACTERISTICS IN RELATION TO ADIPOSITY

6.1. Introduction

Associations between one neighbourhood exposure and an outcome such as obesity might vary according to other neighbourhood factors. Some features of a neighbourhood might matter more (or only) in the presence or absence of other features, e.g. because of synergistic or antagonistic effects.

Building on the results of Chapter 4, where I observed an association between availability of formal physical activity facilities and three measures of adiposity, I consider whether this association is constant regardless of other features of the neighbourhood, or, as seems plausible and perhaps likely, whether the relationship with adiposity varies according to the concurrent level of exposure to fast-food outlets and informal resources for physical activity, such as parks. I hypothesise that among people with limited exposure to parks near home, the role of formal PA facilities will be greater, and therefore the association between their availability and adiposity will be stronger than it is among people with greater local access to parks and other public green and open spaces. I also hypothesise that among people with the highest level of exposure to fast-food outlets near home (measured as proximity, as in Chapter 4), the association between PA facilities and adiposity will be weaker than it is among those who live further from a fast-food outlet, because unhealthy effects of the food environment will dampen the beneficial effects of greater access to PA resources.

I focus on PA facilities as the primary exposure because of the potential for their association with adiposity to be influenced in these two different ways – via alternative resources that might drive energy expenditure, and via an effect on adiposity through energy intake pathways.

At the time of thesis submission I am finalising the paper manuscript in this chapter for submission to *Social Science and Medicine*.

6.2. Research Paper 3

Do neighbourhood characteristics act together to influence BMI? A cross-sectional study of urban parks and takeaway food stores as modifiers of the effect of physical activity facilities

Note: Supplementary material for this research paper is included in Appendix Three.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1510923	**Title**	Ms
First Name(s)	Kate		
Surname/Family Name	Mason		
Thesis Title	Where and for whom does the neighbourhood built environment matter for obesity and health?: Examining sources of effect heterogeneity at multiple scales in the UK adult population		
Primary Supervisor	Neil Pearce		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Social Science and Medicine
Please list the paper's authors in the intended authorship order:	Kate Mason, Neil Pearce, Steven Cummins

Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the analysis with input from SC and NP, and independently undertook the data management, statistical analysis, and writing of the manuscript. SC and NP contributed to the interpretation of results and drafting of the final manuscript.

SECTION E

Student Signature	[redacted]
Date	01/07/2019

Supervisor Signature	[redacted]
Date	01/07/2019

Do neighbourhood characteristics act together to influence BMI? A cross-sectional study of urban parks and takeaway food stores as modifiers of the effect of physical activity facilities

ABSTRACT

Background: Studies exploring associations between neighbourhood environment and obesity often overlook the fact that neighbourhoods are multi-dimensional and that the effects of one environmental exposure may be modified by another. We examine whether associations between neighbourhood availability of formal PA facilities and body mass index (BMI) are modified by neighbourhood availability of public green spaces and proximity to fast-food outlets.

Methods: We used cross-sectional data from the UK Biobank cohort and linked UK Biobank Urban Morphometric Platform (UKBUMP) for 345,254 urban-dwelling adults aged 40-69. We examined associations between objectively measured BMI and the number of formal PA facilities (gyms, pools, etc.) within 1km of each individual's home, testing separately for interactions with the number of local public green spaces, and distance to nearest fast-food/takeaway store. We estimated modifier-stratified associations using multivariable, multilevel regression models to account for a clustered sampling design and potential confounding. Likelihood ratio tests were used to assess statistical interactions.

Results: An inverse association between the number of local PA facilities and BMI was somewhat stronger among people with fewer urban green spaces in their neighbourhood than among those with more green spaces ($P_{interaction}$=0.079). The same relationship between PA facilities and BMI was noticeably attenuated among those who lived closest (<500m) to a fast-food store, compared with people living further away ($P_{interaction}$<0.0001).

Conclusions: Formal PA facilities may buffer against a lack of informal, green resources for PA in areas where the latter are scarce. However, the potential benefits of formal PA facilities in terms of obesity risk may be undermined by an unhealthy food environment close to home. Locating formal PA facilities in places with fewer public green resources and reducing the prevalence of fast-food stores in areas with formal PA resources, may maximise the health benefits to be derived from these neighbourhood resources.

BACKGROUND

Characteristics of neighbourhood environments, such as access to physical activity facilities, green space and fast-food outlets, may be linked to obesity risk. However, the evidence base remains inconsistent for many of these neighbourhood exposures[1–4]. Most studies have tended to focus on exploring associations between single neighbourhood exposures, obesity[5] and obesity-related behaviours[6,7]. One possible explanation for the inconsistencies across studies is that the effects of specific neighbourhood environmental risks may not be universal, but instead vary according to other neighbourhood factors. For example, formal physical activity (PA) facilities are a potentially health-promoting neighbourhood resource[8]. Such facilities (e.g. gyms, swimming pools, sports fields) may play a larger role in areas with fewer informal resources that encourage PA (e.g. parks and other public green space). Conversely, the potentially health-promoting influence of the neighbourhood physical activity environment on energy balance and resulting adiposity may be dampened or overridden by the potentially 'obesogenic' influence of a neighbourhood food environment dominated by fast-food stores. Put another way, neighbourhood characteristics such as the availability of parks, and the proximity of fast food outlets may act as effect-measure modifiers of the relationship between the formal PA environment and obesity.

There is growing recognition that recent increases in obesity prevalence can be viewed as an emergent property of a complex system[9–11] , and it is therefore important to consider any given exposure or risk factor for obesity within its wider context. The presence of effect modification between neighbourhood characteristics is an example of where context might matter – ignoring the underlying distribution of other, effect-modifying neighbourhood characteristics may obscure important effects in some places, and give rise to heterogeneity in findings across different settings. This need to take context into account has been highlighted in a number of recent publications, with respect to population health[12] and the determinants of the major behavioural risks to health[13,14]. Recognising the importance of context and the complexity of obesity and its determinants, it follows that attempts to isolate effects of individual neighbourhood characteristics on health can only ever paint an incomplete picture of how environmental factors influence the health of local residents[15,16].

While there have been many studies in the past decade seeking to unpack some of this complexity in various ways, very few have explicitly examined how multiple dimensions

of the neighbourhood built environment interact with one another such that one neighbourhood characteristic may modify or moderate the effect of another[17]. Instead, efforts have been focussed on characterising overall neighbourhood 'obesogenicity' by combining multiple neighbourhood attributes into a single composite measure[18,19] or using methods such as cluster analysis to identify neighbourhood typologies[20–22]. By understanding whether the effect of one neighbourhood characteristic is modified by the presence of other neighbourhood characteristics, we may better describe how neighbourhoods shape health and behaviour. We may also start to identify settings in which interventions targeting a particular feature of the built environment may have greater (or lesser) potential for reducing or preventing obesity in the populations residing there, and optimise future interventions accordingly[23].

In this paper we focus on potential modification of the relationship between neighbourhood availability of formal PA facilities and adiposity by neighbourhood availability of parks, and neighbourhood fast-food environment. Many formal PA facilities are businesses, and as such they are potentially modifiable via regulatory and commercial levers. Some are run by local authorities, and thus are potentially also amenable to other policy interventions aimed at locating these facilities where they may have the greatest benefit to local populations. In a recent cross-sectional study using UK Biobank, we observed a pattern of lower mean waist circumference, BMI and body fat associated with increasing number of PA facilities in the neighbourhood[24]. Within the same dataset, measures of neighbourhood parks and fast food outlets are available and, as described above, are potential effect modifiers: parks because they can provide alternative opportunities for informal outdoor PA that may be more accessible and appealing than formal PA facilities, and fast-food outlets because unhealthy food environments may negate healthy PA environments. With respect to obesity, the formal PA environment has received less research attention than some other neighbourhood characteristics, particularly in Europe, and findings have been inconsistent[4,8]. Thus, a deeper examination of its relationship with BMI and other measures of adiposity may shed light on settings where intervening on the formal PA environment may be more beneficial, or, alternatively, other modifiable neighbourhood attributes that may boost the potential for local residents to benefit from local formal PA facilities.

We assess these possible environmental effect-modification relationships among adults living in urban residential areas in the United Kingdom by testing the following

hypotheses. First, that the availability of formal PA facilities will be more strongly associated with BMI among people with no parks or other public open/green spaces near their home, than it is among those with greater park availability. Second, that the association between formal PA resources and BMI will be weaker among people who live close to a fast-food outlet, than it is among those who live further from a fast-food outlet.

METHODS

Study design and data collection

We used baseline data from UK Biobank, the scientific rationale, study design and survey methods for which have been described elsewhere[25]. Data were potentially available from 502,656 individuals who visited one of 22 UK Biobank assessment centres across the United Kingdom between 2006 and 2010. Individuals aged 40–69 years living within a 25-mile radius of an assessment centre and listed on National Health Service (NHS) patient registers were invited to participate in the study. The age range was chosen by UK Biobank as an important period for the development of many chronic diseases.

Local environment data

Linked to UK Biobank is a high-resolution spatial database of objectively measured characteristics of the physical environment surrounding each participant's exact residential address known as the UK Biobank Urban Morphometric Platform (UKBUMP). Environmental data in UKBUMP are derived from multiple national spatial datasets using automated processes[26]. The available measures of the local environment include densities of various land uses; proximity to various health-relevant destinations (e.g. GP practices, industrial sites, fast-food outlets); street network accessibility metrics; and attributes of the natural environment. The metrics in UKBUMP were constructed using data collected during the baseline individual assessment phase. No environmental data were collected for the Stockport assessment area, which was the UK Biobank pilot site, leaving 21 assessment areas in scope.

Outcomes

Our primary outcome of interest was Body Mass Index (BMI, kg/m^2). Height and weight measurements were made by trained staff using standard procedures[25] and BMI was centred around its mean and treated as a continuous variable. In sensitivity analyses, waist circumference (in centimetres, measured manually by trained nurses) and body fat

percentage (measured using a bioimpedance machine) were examined as secondary outcomes, to assess the consistency of the results across alternative measures of adiposity.

Primary exposure

Our primary exposure was the neighbourhood availability of formal PA facilities, defined as the total number within a 1000 m street-network buffer around each individual's place of residence (categorised as 0, 1, 2-3, 4 or more, to account for the data being positively skewed). Formal PA facilities were defined at address level as any land use classified in the Commercial-Leisure subcategory of the UK Ordnance Survey AddressBase Premium database. This subcategory comprises a range of indoor and outdoor facilities designed for sporting and leisure activities, such as gyms, swimming pools and playing fields (for details see Supplementary Material, Appendix Three).

Potential effect-measure modifiers

To test hypothesis 1, we examined effect modification by *urban park availability*, measured as the number of parks or other public open/green spaces in a 1000 m street-network buffer around a participant's home address. As with formal PA facilities, this measure is derived from the UK Ordnance Survey AddressBase Premium database. We included any land use categorised as Park; Public Park/Garden; Public Open Space/Nature Reserve; Open Space/Heath/Moorland; or Playground. The distribution of the number of these sites per buffer was highly positively skewed, so for this analysis was categorised as 0, 1, or 2 or more.

To test hypothesis 2, we examined possible effect modification by *fast-food store proximity*, measured by the street-network distance (m) from each individual's residential address to the nearest 'hot/cold fast-food outlet/takeaway', as defined in the UK Ordnance Survey AddressBase Premium database[26]. Distance was then categorised as <500 m, 500-1499 m, or at least 1500 m.

Statistical analysis

We first examined the distributions of each neighbourhood characteristic across the sample by cross-tabulating categories of neighbourhood availability of formal PA facilities with categories of each of the potential modifiers (urban park availability and fast-food store proximity). To test each effect modification hypothesis, we compared multilevel linear models of the independent association between the formal PA environment and

BMI with and without interaction terms for the product of the formal PA environment and each potential modifier. We used likelihood ratio tests to compare the models and we report *p*-values from these tests to indicate the strength of the evidence against the null hypothesis of no effect modification on the additive scale. We then stratified the sample by the potential effect modifier and estimated stratum-specific mean differences (and 95% CIs) in BMI for categories of increasing availability of formal PA facilities, relative to people with no PA facilities within one kilometre of home.

Multilevel models were used to account for the clustering by assessment area in the sampling design, and were estimated with random intercepts and random coefficients for the main exposure, with adjustment for potential confounding by age (years), sex (male/female), self-reported ethnicity (white, south Asian, black, other Asian, mixed white and black, mixed white and Asian, mixed other, or other), highest education level attained (college or university degree; post compulsory education; higher secondary education; secondary education; vocational qualifications; other professional qualification; or none of the above), annual household income (<£18,000, £18,000–30,999, £31,000–51,999, £52,000–100,000, or >£100,000), employment status (paid work, retired, unable to work, unemployed, or other), area deprivation (Townsend score), and neighbourhood residential density (count of residential dwellings within a 1-km street-network buffer of home address, log transformed). Participants without complete covariate data were excluded. The focus of the analysis was to identify moderation or enhancement of the estimated effect of the formal PA environment by other neighbourhood features (specifically parks/fast-food stores). Therefore, we adjusted each model for the set of covariates representing potential confounders of the relationship between the primary exposure and the outcome, and which we identified with the aid of a directed acyclic graph. While adjustment for confounders of the modifier-outcome associations was not essential[27], those sets of confounders are, in this context, likely to be very similar.

We also adjusted each model for the other potential modifier not under examination in that model (i.e. we included fast-food proximity as a covariate in the models testing for effect modification by number of parks, and controlled for number of parks in the model testing for effect modification by fast-food proximity). This made no substantive difference to the point estimates but slightly improved precision of the estimates. In a previous analysis of the association between the formal PA environment and adiposity, we also found that adjustment for diet (total energy intake) did not lead to substantively different

conclusions, but did artificially inflate point estimates due to extensive missing data, therefore we deemed it inappropriate to adjust the models in this paper for diet, particularly as we identified an additional risk of inducing collider bias by adjusting for diet[24].

Missing data and sample restrictions

Perceptions of proximity of food outlets and public amenity of parks and other public open/green spaces are both likely to differ in urban residential areas compared with non-urban areas[28,29]. For example, to a person living in a rural area, many facilities will be relatively far away so proximity to a fast-food store won't mean the same thing as it does to someone living in an urban area if both are measured on the same scale. And in rural areas close to natural landscapes, parks in the immediate neighbourhood may be less important as a potential site of PA than they are for people in the middle of a city. We therefore restricted the analysis to the 86% of the UK Biobank cohort living in areas that are classified by the Office of National Statistics as urban (specifically, where a person's home postcode is located within a city or a town with a population of at least 10,000 people).

Approximately 3% of individuals were missing data on their neighbourhood environment. Data for all other variables were missing at a frequency of <2%, with the exception of income (14.9% missing). The final complete case sample comprised 345,254 individuals.

Sensitivity analyses

To check for consistency across alternative measures of adiposity, we repeated the analyses using waist circumference and body fat percentage rather than BMI. To examine the impact of our exclusion of non-urban participants, we repeated the primary analyses on the full urban and non-urban sample combined, adjusting for urban/non-urban status.

RESULTS

The mean BMI of the analytical sample was 27·5kg/m², and the median number of formal PA facilities within a 1000m street-network distance of participants' homes was two, with just over a quarter of participants having no facilities close to home (Table 6.1). Participants had a median of one park or other public green space within one kilometre of home, and 41% had no parks near home. Median distance to the nearest fast-food store was 1033m, with 20% of the sample living within 500m of their nearest store.

There is a strong positive correlation between number of parks and number of formal PA facilities, with 58% of people who have no formal PA facilities within 1km of home also having no parks close to home, while more than half of those with at least four PA facilities nearby also have at least two parks nearby (Table 6.2). Fast-food proximity is also correlated with access to formal PA facilities, reflecting the clustering of commercial and public services in more densely populated areas. Forty-nine percent of people with no nearby PA facilities also live at least 1500m from their nearest fast-food outlet, compared to only 19% of people with at least four PA facilities living that far from a fast-food outlet (Table 6.2).

Comparison of models with and without a product term for the interaction between formal PA facilities and parks suggests the number of parks within one kilometre of a person's home may weakly modify the association between the formal PA environment and BMI ($P_{interaction}$=0.079). Figure 6.1 shows estimates of the association between formal PA facilities and BMI within strata of urban park availability. As hypothesised, stratification shows that the association between PA facilities and BMI is weak in the areas with the most urban parks, while in contrast there is a clear inverse association between BMI and density of formal PA facilities in the areas with no parks (Figure 6.1). Those living in areas with no parks have a mean BMI 0.21 kg/m² smaller if they have 2-3 formal PA facilities near home, and 0.48kg/m² smaller if they have at least four PA facilities near home, when compared with those with no PA facilities as well as no parks.

Table 6.1 Summary of sample characteristics (N=345,254)

Sample characteristic		
BMI (kg/m2)	Mean (SD)	27.5 (4.8)
Formal PA environment		
Number of facilities in 1km buffer	Median (IQR)	2 (0 - 4)
0	n (%)	94 693 (27.4)
1	n (%)	67 573 (19.6)
2-3	n (%)	86 004 (24.9)
4 or more	n (%)	96 984 (28.1)
Park availability		
Number of parks and other public/open green spaces in 1km buffer	Median (IQR)	1 (0 - 3)
0	n (%)	142 464 (41.3)
1	n (%)	66 713 (19.3)
2 or more	n (%)	136 077 (39.4)
Fast food environment		
Distance to nearest outlet (m)	Median (IQR)	1 033 (577–1 813)
Closer than 500m	n (%)	69 755 (20.2)
500-1499m	n (%)	162 682 (47.1)
At least 1500m	n (%)	112 817 (32.7)
Covariates		
Age	Mean (SD)	56.1 (8.1)
Sex (female)	n (%)	181 887 (52.7)
Ethnicity		
White	n (%)	327 568 (94.9)
South Asian/South Asian British	n (%)	5295 (1.5)
Black/Black British	n (%)	5323 (1.5)
Chinese/other(non-South)Asian	n (%)	2 209 (0.6)
Mixed: White/Black	n (%)	733 (0.2)
Mixed: White/Asian	n (%)	598 (0.2)
Mixed - detail unknown	n (%)	709 (0.2)
Other	n (%)	2819 (0.8)
Income		
Less than 18,000	n (%)	81 547 (23.6)
18,000 to 30,999	n (%)	88 752 (25.7)
31,000 to 51,999	n (%)	89 786 (26.0)
52,000 to 100,000	n (%)	68 045 (19.7)
Greater than 100,000	n (%)	17 124 (5.0)
Education		
College or University degree	n (%)	117 430 (34.0)
A levels/AS levels or equivalent	n (%)	39 523 (11.5)
O levels/GCSEs or equivalent	n (%)	74 020 (21.4)
CSEs or equivalent	n (%)	19 594 (5.7)
NVQ or HND or HNC or equivalent	n (%)	23 405 (6.8)
Other professional qualifications	n (%)	17 389 (5.0)
None of the above	n (%)	53 893 (15.6)
Employment status		
Paid employment or self-employed	n (%)	210 587 (61.0)
Retired	n (%)	106 497 (30.9)
Unable to work	n (%)	10 785 (3.1)
Unemployed	n (%)	5340 (1.6)
Home duties/carer/student/volunteer/other	n (%)	12 045 (3.5)
Area deprivation (2001 Townsend index)	Median (IQR)	-2.0 (-3.6 - 0.7)
Residential density (residential addresses per 1km street network buffer)	Median (IQR)	2152 (1352–3344)

Table 6.2 Bivariate associations between neighbourhood characteristics

	Availability of parks and other open/green spaces							
Formal PA facilities	No parks		One park		≥2 parks		Total	
	n	%	n	%	n	%	n	%
0	54513	57.6	16737	17.7	23443	24.8	94693	100.0
1	31683	46.9	13000	19.2	22890	33.9	67573	100.0
2-3	32031	37.2	17925	20.8	36048	41.9	86004	100.0
4 or more	24237	25.0	19051	19.6	53696	55.4	96984	100.0
Total	*142464*	*41.3*	*66713*	*19.3*	*136077*	*39.4*	*345254*	*100.0*
	Distance to nearest fast-food store							
Formal PA facilities	<500m		500-1499m		At least 1500m		Total	
	n	%	n	%	n	%	n	%
0	8338	8.8	39710	41.9	46645	49.3	94693	100.0
1	10126	15.0	32573	48.2	24874	36.8	67573	100.0
2-3	18762	21.8	43964	51.1	23278	27.1	86004	100.0
4 or more	32529	33.5	46435	47.9	18020	18.6	96984	100.0
Total	*69755*	*20.2*	*162682*	*47.1*	*112817*	*32.7*	*345254*	*100.0*

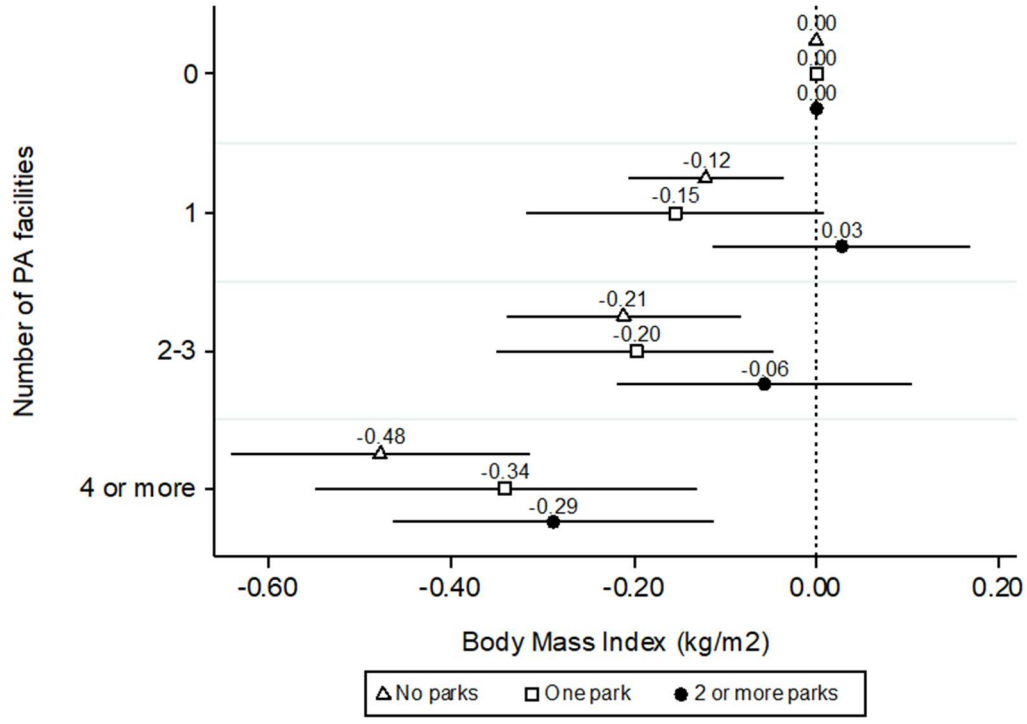


Figure 6.1 Association between number of formal PA facilities and BMI, stratified by park availability

Figure shows park availability-stratified, fully adjusted mean differences in BMI and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and fast-food store proximity.

There is strong statistical evidence that the PA environment-BMI association is modified by proximity to a fast-food store ($P_{interaction}<0.0001$). In line with our second hypothesis, stratified results showed that among people living within 500m of a fast-food/takeaway store, the association between density of nearby formal PA facilities and BMI is considerably less apparent than it is among those who live further from a fast-food store (Figure 6.2).

Results were broadly consistent across alternative measures of adiposity: the same patterns of effect modification we observed for BMI were also present for waist circumference and body fat percentage (see Supplementary Material, Appendix Three). When we included respondents living in non-urban areas, and adjusted for urban/non-urban status, the patterns across stratum-specific models mirrored those observed in the urban-only sample, but statistical evidence of an interaction with park availability was weaker (see Supplementary Material, Appendix Three).

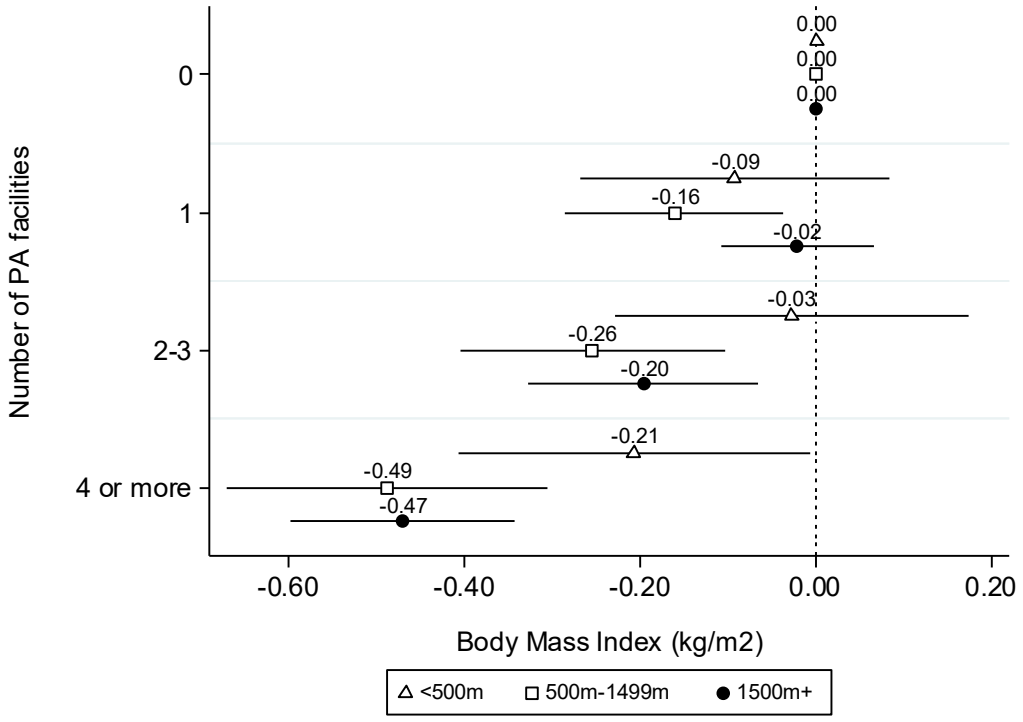


Figure 6.2 Association between number of formal PA facilities and BMI, stratified by distance to nearest fast-food/takeaway store

Figure shows fast-food proximity-stratified, fully adjusted mean differences in BMI and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and park availability.

DISCUSSION

This study provides evidence to support the presence of environmental effect modification in this large sample of mid-life adults from across the UK. In stratified models we observed that the relationship between access to formal PA facilities and BMI is much weaker among people living close to a fast-food store than it is among those living further away from such a store. The association between the formal PA environment and BMI is also somewhat stronger among people living in areas with fewer urban parks and other public open/green spaces than it is among people living in areas with more of these, where there is likely to be greater opportunity for informal, outdoor PA.

These findings suggest that locating formal PA facilities close to residential areas has potential to reduce BMI among local residents, but that other contextual features of the neighbourhood are likely to influence these potential benefits. While we cannot infer causality from this cross-sectional study, our results suggest that increasing the availability of formal PA facilities may have the most potential to reduce population obesity in areas that have the lowest densities of parks and, in particular, least exposure to fast-food stores. This is consistent with the hypothesis that in areas with fewer parks and other green spaces, formal PA facilities provide valuable opportunities for PA that are otherwise lacking. Meanwhile, formal PA facilities may have limited influence in areas with fast-food stores close to people's homes – even if they do serve to increase PA there, our findings suggest the positive benefits for body weight may be dampened by the influence of an unhealthy food environment. As a public health intervention, the introduction of PA resources such as gyms, swimming pools and other sports facilities in the neighbourhoods of people in close proximity to fast-food stores may be ineffective unless coupled with interventions aimed at minimising the influence of fast-food stores. In urban areas well served by parks, interventions involving formal PA facilities may not be a priority and a focus on other environmental interventions may be more effective in improving population health. More causally focused study designs are needed to confirm these implications.

Our findings also highlight the possibility that effect heterogeneity patterns such as these may also apply to relationships between other neighbourhood characteristics and a range of health outcomes. Such heterogeneity may, at least partially, explain inconsistent results across studies and settings. We examined only three neighbourhood characteristics,

motivated by two plausible effect modification hypotheses. Similar interactions may also exist between other neighbourhood characteristics.

PA environment effects on obesity in the UK have sometimes been shown to be concentrated in individuals with higher incomes[24] or more education[30]. While we have not examined an additional interplay with socioeconomic status in this paper, it is likely that an important caveat applies to our findings: that the potential for the benefits of local access to formal PA facilities to be maximised via a supportive broader neighbourhood environment relies on PA facilities being accessible and affordable for all.

Strengths and limitations

These findings are a novel contribution to this area of research. To our knowledge, no other studies in the UK, and few outside it, have explicitly examined modification of the association between the formal neighbourhood PA environment and adiposity (or any other obesity-related outcome) by other neighbourhood built environment characteristics. One similar study in the United States concluded that combined changes to the food and PA environments would have stronger and more consistent effects on BMI than changes that addressed only one dimension or the other. Our findings provide similar evidence in a European context[17]. Others have examined composite measures of neighbourhood obesogenicity or other similar constructs, and while such research importantly recognises and draws attention to the complex and multidimensional nature of neighbourhood environments, it lends itself to more general conclusions about the importance of holistic healthy urban planning, rather than moving towards specific policy recommendations. Furthermore, these studies typically rely on data-driven approaches such as latent class analysis, and this makes generalising to other populations challenging. Here, although our findings require confirmation using longitudinal data and more causally focussed methods, and could be tested by evaluating real-world interventions in different places, the results provide evidence in support of two clearly defined and theoretically grounded effect modification hypotheses, and point to prioritisation of built environment interventions that take into account local context. Our findings were also consistent across multiple adiposity measures, and robust to various modelling choices, as shown in the sensitivity analyses we performed.

UK Biobank is a very large and geographically diverse cohort, providing unique opportunities in this field of research. However, the sample is based on only a 5.5%

response rate, and does show some evidence of 'healthy volunteer' bias[31]. A further potential source of selection bias is the exclusion of a large number of observations without data on household income. Due to the cross-sectional nature of this study it is impossible to draw strictly causal inferences about these patterns of association, but the results do lend support to two a priori hypotheses about plausible interactive effects of multiple aspects of the neighbourhood environment. There remains the possibility that the observed main associations are driven by people with lower BMIs self-selecting into 'healthier' neighbourhoods. Studies that have directly examined the influence of self-selection on neighbourhood-health effects have reached inconsistent conclusions about the likely bias this may induce[32–34]. For those fortunate enough to have substantial choice over where they live, the presence of formal PA facilities alone is unlikely to be a major governing factor in that choice, but the presence of such facilities is likely to coincide with other facilities that may enhance the desirability of a neighbourhood, including parks and other green space (as we observed in Table 6.2). That said, in this sample the individuals living in neighbourhoods with high densities of parks and PA facilities are, contrary to expectation, not necessarily those with the highest incomes, or living in the least deprived postcodes in the study.

Due to the size of the sample and the breadth of the neighbourhood characterisation, large-scale automated processes were used to derive the environmental metrics on which we have relied here[26], and while the best readily available for conducting these analyses at scale, those metrics are of varying quality, accuracy and suitability. It should be noted that the fast-food proximity measure in particular may be susceptible to some misclassification error, and that the measure of park availability does not account for the quality of those spaces. There is also a possibility that if any of the main associations are confounded in one stratum of the potential effect modifier and not another, we may erroneously infer effect modification when none is present[35].

These findings therefore provide preliminary observational evidence for plausible interplay between multiple aspects of the built environment in the UK, but further research using more causally focussed approaches such as longitudinal or quasi-experimental study designs is needed. An additional implication of our findings is that evaluations of PA environment interventions, particularly those pertaining to formal PA facilities, may be underestimating the impact of the intervention if possible moderation by local park availability and food environments is ignored.

Residential neighbourhoods are complex and multidimensional and the examination of the effect of individual environmental characteristics on obesity in isolation overlooks this complexity. Here we examined whether some neighbourhood characteristics modify the effect of others to better understand how they may operate in concert to influence BMI. Our findings suggest that formal PA facilities may buffer against a lack of informal, green resources for PA such as parks, in areas where the latter are scarce, but that potential benefits of formal PA facilities in terms of adiposity may be undermined by the presence of fast-food stores close to home. Reducing the prevalence of fast-food outlets in areas with formal PA resources, and prioritising the location of formal PA facilities in places without public parks, may maximise the potential for PA facilities to influence adiposity. An approach to urban planning that takes into account moderating effects of other neighbourhood characteristics is required in order to maximise the population health benefits of the urban built environment.

REFERENCES

1 Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity* 2015; **23**: 1331–44.

2 Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: a systematic review of naturally occurring experiments. *Obes Rev* 2015; **16**: 362–75.

3 Lachowycz K, Jones AP. Greenspace and obesity: a systematic review of the evidence. *Obes Rev* 2011; **12**: e183–9.

4 Mackenbach JD, Rutter H, Compernolle S, *et al.* Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. *BMC Public Health* 2014; **14**: 233.

5 Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 2010; **16**: 175–90.

6 Caspi CE, Sorensen G, Subramanian S V, Kawachi I. The local food environment and diet: a systematic review. *Health Place* 2012; **18**: 1172–87.

7 Van Holle V, Deforche B, Van Cauwenberg J, *et al.* Relationship between the physical environment and different domains of physical activity in European adults: a systematic review. *BMC Public Health* 2012; **12**: 807.

8 Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation* 2012; **125**: 729–37.

9 Butland B, Jebb S, Kopelman P, *et al.* Foresight Tackling Obesities: Future Choices – Project report. London, UK, 2007

10 Swinburn BA, Sacks G, Hall KD, *et al.* The global obesity pandemic: shaped by global drivers and local environments. *Lancet* 2011; **378**: 804–14.

11 Rutter H, Savona N, Glonti K, *et al.* The need for a complex systems model of evidence for public health. *Lancet* 2017; **390**: 2602–4.

12 Craig P, Di Ruggiero E, Frohlich KL, Mykhalovskiy E and White M, on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. Taking account of context in population health intervention research: guidance for producers, users and funders of research. Southampton: NIHR Evaluation, Trials and Studies Coordinating Centre; 2018.

13 Watts P, Phillips G, Petticrew M, Harden A, Renton A. The influence of environmental factors on the generalisability of public health research evidence: physical activity as a worked example. *Int J Behav Nutr Phys Act* 2011; **8**: 128.

14 Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. *Int J Behav Nutr Phys Act* 2017; **14**: 156.

15 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev* 2011; **12**: e95–106.

16 Myers CA, Denstel KD, Broyles ST. The context of context: examining the associations between healthy and unhealthy measures of neighborhood food, physical activity, and social environments. *Prev Med* 2016; **93**: 21–6.

17 Boone-Heinonen J, Diez-Roux A V., Goff DC, *et al.* The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study. *PLoS One* 2013; **8:** e85141.

18 Tseng M, Thornton LE, Lamb KE, Ball K, Crawford D. Is neighbourhood obesogenicity associated with body mass index in women? Application of an obesogenicity index in socioeconomically disadvantaged neighbourhoods. *Health Place* 2014; **30**: 20–7.

19 Saelens BE, Sallis JF, Frank LD, *et al.* Obesogenic neighborhood environments, child and parent obesity: The neighborhood impact on kids study. *Am J Prev Med* 2012; **42**: e57–64.

20 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Neighbourhood typologies and associations with body mass index and obesity: A cross-sectional study. *Prev Med* 2018; **111**: 351–7.

21 Adams M, Ding D, Sallis JF, *et al.* Patterns of neighborhood environment attributes related to physical activity across 11 countries: a latent class analysis. *Int J Behav Nutr Phys Act* 2013; **10**: 34.

22 Meyer KA, Boone-Heinonen J, Duffey KJ, *et al.* Combined measure of neighborhood food and physical activity environments and weight-related outcomes: the CARDIA study. *Heal Place* 2015; **33**: 9–18.

23 Economos CD, Hatfield DP, King AC, Ayala GX, Ann Pentz M. Food and physical activity environments: an energy balance approach for research and practice. *Am J Prev Med* 2015; **48**: 620–9.

24 Mason KE, Pearce N, Cummins S. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *Lancet Public Heal* 2018; **3**: e24–33.

25 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007.

26 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP): a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21**: 135–48.

27 Vanderweele T. Explanation in Causal Inference: methods for mediation and interaction. New York : Oxford University Press, 2015.

28 Dennis M, James P. Evaluating the relative influence on population health of domestic gardens and green space along a rural-urban gradient. *Landsc Urban Plan* 2017; **157**: 343–51.

29 Maas J, Verheij RA, Groenewegen PP, De Vries S, Spreeuwenberg P. Green space, urbanity, and health: how strong is the relation? *J Epidemiol Community Health* 2006; **60**: 587–92.

30 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Associations between the combined physical activity environment, socioeconomic status, and obesity: a cross-sectional study. *Perspect Public Health* 2018; **138**: 169–72.

31 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**: 1026–1034.

32 James P, Hart JE, Arcaya MC, Feskanich D, Laden F, Subramanian S V. Neighborhood self-selection: the role of pre-move health factors on the built and socioeconomic environment. *Int J Environ Res Public Health* 2015; **12**: 12489–504.

33 McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int J Behav Nutr Phys Act* 2011; **8**: 125.

34 Grafova IB, Freedman VA, Lurie N, Kumar R, Rogowski J. The difference-in-difference method: assessing the selection bias in the effects of neighborhood environment on health. *Econ Hum Biol* 2014; **13**: 20–33.

35 Kamangar F. Effect modification in epidemiology and medicine. *Arch Iran Med* 2012; **15**: 575–82.

Chapter 7. GEOGRAPHICAL HETEROGENEITY IN ASSOCIATIONS BETWEEN NEIGHBOURHOOD BUILT ENVIRONMENTS AND ADIPOSITY

7.1. Introduction

Having observed in Chapter 6 that the extent to which one neighbourhood characteristic matters for adiposity depends on other characteristics of the neighbourhood, we might also conceive of similar types of modifying, contextual factors operating at other geographical scales. Rather than assuming effects of the neighbourhood environment apply uniformly across space, we might anticipate that the magnitude of neighbourhood effects will vary geographically, and might be moderated by characteristics of the wider contexts in which neighbourhoods are located.

In this next chapter, I first examine whether the associations between two neighbourhood characteristics (availability of PA facilities and proximity to fast food) and adiposity (specifically BMI) vary geographically across Local Authorities of England. Evidence of geographical heterogeneity across a large area within a single study would suggest this as one possible explanation for inconsistent results from studies in different settings. I then investigate whether any such geographical heterogeneity observed is explained by locally varying, macro-environmental factors.

I focus on two dimensions of the neighbourhood environment, and Local Authorities as the geographical unit by which their relationships with BMI might vary. These particular relationships can be considered as case studies of sorts. It is possible to conceive of any neighbourhood characteristic (or indeed some other exposure) having a differential effect on health depending on *where* you look – be that from one local authority or city to the next, one country to another, etc. And while administrative units such as Local Authorities, countries, etc. are in some sense arbitrary, they may be meaningful if local policies or cultural factors are at least roughly determined or delineated by these boundaries. For the purposes of analysis, administrative units also provide ready access to data on potentially relevant area-level modifiers (as I demonstrate using two such variables).

Through the examination of geographical heterogeneity in neighbourhood-health relationships, and possible drivers of that heterogeneity, this chapter uses observational data to highlight the need for an awareness of context and the importance of context-

specific adaptation of built environment interventions for health. I am preparing the paper manuscript in this chapter for submission to *Health and Place*.

7.2. Research Paper 4

Geographical heterogeneity in associations between the neighbourhood built environment and BMI

Note: Supplementary material for this research paper is included in Appendix Four.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1510923	**Title**	Ms
First Name(s)	Kate		
Surname/Family Name	Mason		
Thesis Title	Where and for whom does the neighbourhood built environment matter for obesity and health?: Examining sources of effect heterogeneity at multiple scales in the UK adult population		
Primary Supervisor	Neil Pearce		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Health and Place
Please list the paper's authors in the intended authorship order:	Kate Mason, Neil Pearce, Steven Cummins

Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the analysis with input from SC, and undertook the data management, sourcing of external datasets, statistical analysis, and writing of the manuscript. SC and NP contributed to the interpretation of results and drafting of the manuscript.

SECTION E

Student Signature	[redacted]
Date	01/07/2019

Supervisor Signature	[redacted]
Date	1/07/2019

Geographical heterogeneity in associations between the neighbourhood built environment and BMI

ABSTRACT

Background: The effects of residential neighbourhood environments on health may vary across geographical space, and differences in local contexts could influence how much a given neighbourhood characteristic matters for the health of local residents.

Methods: Linking UK Biobank data from 302,952 urban-dwelling adults in England to publicly available Local Authority-level data, we examine (a) whether associations between BMI and two characteristics of the neighbourhood built environment (availability of formal physical activity facilities, and fast-food proximity) vary by Local Authority District, and (b) whether cross-level interactions with local authority-level physical features (natural landcover) and socio-cultural attributes (local descriptive obesity norms) reveal evidence of effect modification by these features of the wider contexts in which neighbourhoods are located.

Results: We found that the relationship between availability of neighbourhood physical activity facilities and BMI does vary between Local Authority Districts across urban England. Though differences were small, there was some evidence that the association was stronger among people living in areas with less natural landcover and potentially therefore a greater reliance on/normalisation of the use formal physical activity facilities, especially in areas outside of London. We also found that the relationship between proximity of fast-food stores to people's homes and BMI varied geographically across England. There was little evidence that local descriptive obesity norms is an important modifier of this association.

Conclusions: This paper highlights the importance of considering potential geographical heterogeneity in relationships between the built environment and health, and the implications for generalisability of research findings. By seeking to understand sources of heterogeneity, we may be able to better tailor and target built environment interventions for health.

BACKGROUND

The built environment in residential neighbourhoods can affect the weight status of the residents who live there, by influencing diet and physical activity (PA) behaviours. The makeup of the retail food environment, availability of places to engage in recreational physical activity, and how 'walkable' a neighbourhood is, have all been linked to diet or PA, and to obesity risk[1–3]. However, recent systematic reviews of the literature on the neighbourhood environment and obesity have concluded that, despite a wealth of research, the current body of evidence "does not tell a clear story"[4] and "does not allow robust identification of ways in which [the] physical environment influences adult weight status"[5]. One possible reason for the inconsistency of the evidence base is that neighbourhood effects may not be uniform across geographical space. Neighbourhood effects may be stronger in some places than others, and particular characteristics of a neighbourhood may have more or less influence depending on features of the broader local context.

1.1 Geographical heterogeneity in the evidence for relationships between neighbourhood environments and health

There is growing evidence that relationships between residential neighbourhood characteristics and obesity are stronger is some settings than in others. While it has been noted that relationships between food environment and obesity appear to be stronger in North America than in other settings[6,7], less well recognised is the fact that very mixed findings are observed even within a single region or country. Considering the influence of the PA environment, for example, a recent review reported that across European studies, evidence for the influence of parks and PA facilities and their association with obesity is too mixed to draw conclusions[5]. Even within the UK the evidence is inconsistent: some recent studies have found that local access to recreation facilities is negatively associated with adiposity[8], and with obesity[9], while others have found no association between access to parks and PA facilities and either obesity or change in BMI over time[10,11]. The same is true of the food environment and obesity in the UK: studies in London[12], Leicester[13], Cambridgeshire[14,15] and Norfolk[16] found greater exposure to fast-food outlets was associated with higher BMI or greater odds of obesity, yet studies in the North East of England[17] and in Leeds[18,19] showed no such association. In the United States, two recent studies of cross-sectional relationships between objective and perceived measures of neighbourhood built environments and BMI across the US found that significant geographical variation existed[20,21]. The authors of those studies concluded that this may

explain why inconsistent findings often emerge across studies in single geographical areas, and why built environment interventions do not consistently work in reducing population obesity.

1.2 Contextual influences on health, operating at multiple scales

The residential neighbourhood can be defined in various ways[22] but broadly refers to the local area in which a person lives. Neighbourhoods are themselves nested within wider geographical and administrative settings (cities, counties, nations, etc.), and just as neighbourhood characteristics may influence obesity-related health behaviours and health outcomes, so too can physical, political, economic, and socio-cultural factors operating at the macro-environmental scale of those larger units within which neighbourhoods are nested[23]. Such factors may include quality of local government and public sector expenditure[24,25]; climate and weather[26,27]; economic prosperity[28,29]; greenspace[30]; and social norms regarding health behaviours and obesity[31,32].

With factors operating at multiple levels to influence health, macro-environmental attributes of the larger geographical units in which neighbourhoods are nested are potential modifiers of more local neighbourhood effects on health[33]. Variation in macro-environmental factors may explain some of the observed heterogeneity in the magnitude of neighbourhood-health associations from one study setting to another. Although conceptual models recognising these complex and multilevel relationships have existed for some time[23,34], the potentially modifying roles of wider contextual factors are typically ignored. Most studies of associations between neighbourhood environments and health assume – implicitly, at least – that neighbourhood effects are both uniform across space and potentially generalisable to other settings. It is plausible, however, that variation in wider contextual factors undermines both these assumptions. This may explain the abundance of inconsistent findings from studies conducted in different settings, including different parts of the same country.

There have been calls in recent years to recognise and empirically examine likely modification of built environment health effects[35,36], partly in response to observed inconsistency of findings, partly driven by theory, and increasingly made possible by larger sample sizes. As yet, very few studies have examined whether and how neighbourhood-obesity associations vary geographically or according to explicitly place-based, macro-environmental variables.

One way to examine both the presence and correlates of possible geographical heterogeneity in relationships between neighbourhood characteristics and health outcomes such as obesity, is to conduct studies with broad geographical coverage spanning a wider spectrum of contexts, for example across multiple cities within one country. In contrast with meta-analytical approaches, which have been hampered by the substantial methodological heterogeneity of existing studies[37,38], this approach allows explicit comparison of effect estimates across different areas within the same study, while holding constant methods that may otherwise vary across separate studies and make comparison difficult. This approach also provides an opportunity to examine interactions with variables at multiple scales other than the individual or the neighbourhood, potentially providing insights into the interplay of health determinants at multiple scales.

If patterns of association are similar across space, then results of other studies of that relationship are likely to be broadly generalisable from one setting to another. Taken a step further, findings from natural experiments or intervention studies might also be assumed to be transferable to similar populations in other settings if the relationship is stable across geographical space. On the other hand, heterogeneity in associations from place to place would undermine the generalisability of findings from studies with narrow geographical coverage. Furthermore, if heterogeneous effects are driven by attributes of some larger areal unit of analysis, then an understanding of such effect modification would ultimately be important for informing the tailoring and targeting of interventions based on local context[39].

In this chapter I provide two worked examples of how an association between a neighbourhood characteristic and adiposity might vary geographically and how this might be partly explained by locally varying macro-environmental effect modifiers.

1.3 Example 1: neighbourhood availability of formal PA facilities and BMI

As noted above, previous studies in the UK have examined the association between the availability of PA facilities close to home and obesity-related measures such as BMI, with some inconsistent findings. Taking that relationship as our first example of a neighbourhood-health relationship, we hypothesise that the relationship between neighbourhood availability of PA facilities and BMI varies geographically across the country. If that is the case, it may arise because macro-environmental factors operating at a sub-national scale within which neighbourhoods are nested may modify the relationship. Such factors could be socio-cultural or economic in nature, or may reflect features of the physical landscape or climate. In this first example we focus on a potential modifier from

the physical landscape. Residents of cities and towns surrounded by a lot of natural landcover (woodland, moors, beaches, etc.) have enhanced opportunities for outdoor, informal PA even if those natural spaces are not within one's immediate neighbourhood. This increased exposure to natural landcover may also contribute to a local culture of outdoor recreation. In such places, a weaker reliance on or normalisation of using formal PA facilities such as gyms and leisure centres close to home may exist, reducing the magnitude of association between the neighbourhood availability of these facilities and BMI.

1.4 Example 2: fast-food proximity and BMI

A relationship with BMI has also been demonstrated in some but not all studies of exposure to fast-food outlets. We previously identified a weak association between proximity of home address to nearest fast-food/takeaway store in the UK Biobank cohort, while findings from smaller and geographically narrower samples in various settings across the UK have yielded inconsistent results. It may be that our weak association overall masked localised heterogeneity in the magnitude of the association. Such geographical variation might contribute to inconsistent findings from studies in different settings. In our second example we therefore test the hypothesis that the association between the proximity of fast-food outlets to people's homes and BMI varies geographically across the country. And, just as for the earlier example of PA facilities, if such heterogeneity exists, it may arise as a result of effect modification by locally varying macro-environmental factors. In this second example, we consider a socio-cultural attribute of the macro environment as a potential modifier. Spatial variation in the prevalence of particular traits (e.g. obesity) or behaviours (e.g. diet) creates what are known as local descriptive social norms[40]. Theoretically, in areas where obesity is 'normalised' due to a high prevalence of obesity, the influence of unhealthy food environments on BMI may be unfettered by social pressure to be a healthy weight. In contrast, where obesity prevalence is lower, we would expect stronger social pressure to maintain a healthy weight, and such pressure may act as a counter to easy access to fast food, thereby attenuating the main association. We therefore test the hypothesis that the association between the fast-food environment and BMI is weaker in local authorities where there is a lower prevalence of adult obesity.

Macro-environmental factors may operate at various scales, and for the purposes of analysis the choice of scale is important. One potentially relevant scale is the scale at which local government is organised. With respect to our examples, planning regulations and resource allocation decisions determined by local government influence the local authority area as a whole (e.g. local authorities often contribute to the management of

natural areas in the UK, and local public health teams and others in the LA enact various strategies to curb obesity) and on the neighbourhoods located within that area (e.g. UK local authorities provide funding for local PA facilities, and regulate planning decisions about the food environment). While the boundaries of such areas are somewhat arbitrary, they nonetheless define the scale at which many health-relevant decisions are made. They are also a scale at which considerable data are collected, which may be used to inform decision making. Consequently, understanding if and how factors operating at the local government level modify neighbourhood-health relationships could be important for informing local planning decisions and targeting built environment interventions more effectively; for instance, improving access to PA facilities in contexts where they are expected to have greater influence, while focusing efforts on improving neighbourhood food environments in settings where PA environments have less influence on BMI.

The principles underlying the two worked examples we provide here may also apply in general terms to other health-relevant neighbourhood exposures and other macro-environmental modifiers.

1.4 Study aims

We make use of a very large and geographically diverse sample of mid-aged adults from the UK Biobank cohort to examine whether the relationship between (a) the neighbourhood PA environment and BMI, and (b) neighbourhood fast-food proximity and BMI, vary between local authority districts across England. We use local authority district (LAD) boundaries to delineate the wider context within which neighbourhoods are nested. LADs are the 326 sub-national units of local governance in England. For each of the two associations between the neighbourhood characteristics and BMI, we explore potential effect modification by a different attribute of the wider LAD context, as a demonstration of how physical and socio-cultural macro-environmental factors might interact with neighbourhood factors to influence health. For the neighbourhood PA environment, we examine the potential modifying role of the percentage of land cover classified as 'natural' in the surrounding LAD, and for the fast-food environment, we explore the potential modifying role of local descriptive obesity norms in the LAD, represented by adult obesity prevalence.

METHODS

2.1 UK Biobank and UK Biobank Urban Morphometric Platform

We used baseline data from UK Biobank, the details of which are reported elsewhere[41]. Briefly, 502,656 adults aged 40-69 and registered with the National Health Service (NHS) were recruited from 25-mile radius assessment areas in 22 locations across England, Scotland and Wales, and underwent detailed baseline assessment spanning health, lifestyle, demographic and socioeconomic characteristics, between 2006 and 2010. Linked to UK Biobank via the home address of each participant is the UK Biobank Urban Morphometric Platform (UKBUMP), a high-resolution spatial database of a wide range of objectively measured characteristics of the physical environment surrounding each individuals' residential address, derived from multiple national spatial datasets[42]. Local environment metrics include, among others, the densities of various land use types, and street-network distances to health-relevant destinations, both derived from the Ordnance Survey AddressBase Premium database. We used the land-use densities data in UKBUMP to derive a measure of neighbourhood availability of formal PA facilities, and the distance-to-nearest-destination data to derive a measure of proximity to a takeaway/fast-food outlet.

2.2 Outcome: Body Mass Index

Body Mass Index (BMI, kg/m^2) was calculated from weight and height measurements made by trained staff using standard procedures[41]. BMI was normally distributed and treated as a continuous outcome variable.

2.3 Exposure 1: Availability of PA facilities

Availability of PA facilities was operationalised as the number of formal PA facilities within a one-kilometre street-network distance of a person's home. These facilities included gyms, swimming pools, leisure centres, playing fields and others detailed in Supplementary Table 1. The measure was included in models as a continuous variable, to enable estimation of a single coefficient for each LAD and visual representation of these to display geographical heterogeneity. It also had the additional benefit of allowing more parsimonious random effects models in the second stage of the analysis, as only a single random effect for the exposure was required. Assuming a linear relationship with BMI is consistent with results of our previous analyses using these data, where a categorical operationalisation of the exposure was found to have an approximately linear relationship

with BMI. Due to the highly positively skewed distribution of this variable, we top-coded the number of facilities at 15.

2.4 Exposure 2: Fast-food proximity

Proximity to a fast-food store was defined as the street-network distance (metres) from each individual's residential address to the nearest 'hot/cold fast-food outlet/takeaway', as defined in the UK Ordnance Survey AddressBase Premium database [42]. Distances were $\log_{10}$-transformed for ease of interpretation, so that a one-unit increase represented a 10-fold increase in distance to the nearest outlet (e.g. 100m to 1000m).

2.4 Assignment to Local Authority Districts and linkage of effect modifier data

The UKBUMP local environment metrics are based on exact home address locations and then linked to the UK Biobank cohort and made available to approved researchers. Due to privacy restrictions, the exact address coordinates of participants are not themselves routinely available to researchers; instead approximate coordinates (rounded to the nearest 1 km) are available. We used these approximate coordinates to geocode participants and assign them to the LAD in which they reside, using QGIS v2.14 (2016). We identified 91 address points that appeared to be incorrect because they were well outside the geographical scope of the UK Biobank study, and excluded these from this analysis.

Following the assignment of participants to LADs, we undertook additional linkage based on the LAD boundaries to three external, publicly available data sources. As administrative units, LADs are well described in publicly available datasets spanning multiple domains, enabling us to obtain the following LAD-level variables for analysis: (1) percentage of land cover in the LAD classified as 'natural' based on Corine Land Cover data from 2012[43], as compiled in the Land Cover Atlas of the UK[44]; (2) estimated adult obesity prevalence in 2003-05 derived from the Health Survey for England[45]; and (3) gross disposable household income (GDHI) per capita for 2006[46]. Natural land cover was examined as a potential modifier of the effect of the availability of PA facilities, obesity prevalence was examined as a potential modifier of the effect of proximity to a fast-food outlet, and GDHI was included as a possible confounding variable in the multilevel analyses for both exposures.

The 'natural' land cover definition is based on the Corine Land Cover classifications and includes all land cover that is neither 'artificial' (urban, industrial, commercial, transport, mining etc) nor 'agricultural'. The 'natural' classification spans land cover types such as forests, grasslands, moorland, beaches, wetlands, and water bodies. It does not include farmland such as pastures, which is classified as 'agricultural' or urban green areas such as

parks and sport and leisure facilities, which are classified in Corine as 'artificial' and in the Land Cover Atlas of the UK as their own category of 'urban green'. Corine Land Cover data are accurate to approximately 25 metres. Natural landcover percentage was positively skewed so it was square-root transformed prior to analysis. As people living in rural areas may have a different relationship to the natural environment[47,48], we restricted the analysis to the 86% of the UK Biobank cohort living in areas that are classified by the Office of National Statistics as urban (specifically, where a person's home postcode is located within a city or a town that has a population of at least 10,000 people).

Obesity prevalence estimates were only available for LADs in England, so we restricted all our analysis in this paper to UK Biobank participants residing in England. This also reduced the risk of confounding due to contextual differences that might arise from historical or current differences between the devolved nations of the UK.

2.5 Statistical analysis

For the primary association between each of the two neighbourhood exposures and BMI, we estimated a separate linear regression model for each LAD, with robust standard errors. Models were adjusted for potential confounding by age (years), sex (male/female), highest education level attained (Degree; A level or equivalent; O level or equivalent; CSE or equivalent; NVQ/HND/HNC; other professional qualification; none of the above), annual household income (<£18,000; £18,000-£30,999; £31,000-£51,999; £52,000-£100,000; >£100,000), employment status (paid work, retired, unable to work, unemployed, other), area deprivation (Townsend score), and neighbourhood residential density (count of residential features within a 1km street-network buffer of home address, log transformed). Residential density has been shown to be associated with obesity-related outcomes[35], and may also serve as a proxy for the density or proximity of other neighbourhood resources that will be correlated with the exposures of interest. Models of the availability of PA facilities were also adjusted for fast-food proximity, while models of fast-food proximity were adjusted for availability of PA facilities. We excluded 30 LADs with fewer than 200 study participants (1006 observations in total) to avoid estimating LAD-stratified effects based on small numbers of people in an area. The LAD-specific estimates were plotted, and also mapped using QGIS to visualise the geographic variation in the estimated association.

Prior to testing for interactions with LAD-level variables we calculated the overall proportion of variation in BMI that was attributable to differences between local authorities rather than within-LA differences between individuals. This was done by

estimating a random intercept model clustered at the level of Local Authority, but without the inclusion of any local authority variables in the model, and using the *estat icc* postestimation command to estimate the variance partition coefficient (VPC).

To examine cross-level interactions between each neighbourhood exposure and our selected attributes of the wider LAD context, we used multilevel models with random intercepts and random effects allowing the association to vary by Local Authority, and interaction terms between the exposure and potential modifier. These models were adjusted for the same covariates as the single-level LAD-specific models, plus LAD-level gross household disposable income per capita to control for possible confounding effects of the wider socioeconomic context. The exposure variables were cluster-mean centred[49] so that the effect estimates represent the mean difference in BMI for each unit change in the exposure relative to the LAD mean of the exposure. While the effect modifiers were modelled as continuous variables, results of the regression models were plotted to show mean BMI difference per unit change in the exposure according to tertile of the effect modifier, to aid visualisation.

2.6 Sensitivity analysis

It may be that London exerts a strong influence over the nation-wide model of cross-level interactions, so we repeated that stage of each analysis on a sample that excluded the 20 Local Authorities represented in UK Biobank located in London.

2.7 Ethics

UK Biobank has ethics approval from the North West Multi-centre Research Ethics Committee (reference 16/NW/0274), the Patient Information Advisory Group (PIAG), and the Community Health Index Advisory Group (CHIAG). Additional institutional ethics approval was granted to this particular study by the London School of Hygiene and Tropical Medicine's Research Ethics Committee in September 2016 (reference 11897).

RESULTS

3.1 Descriptive statistics

The complete case sample used in this analysis was made up of 302,952 UK Biobank participants from 122 of the 326 local authority districts in England. Across the individual-level sample, the median number of PA facilities in a person's neighbourhood was two, the median distance to nearest fast-food/takeaway store was 996 metres, and the mean BMI was 27.5 kg/m^2. Across the 122 LADs, the percent of land cover classified as 'natural' ranged

from zero to 49.7% and the median value was 4.9%. The majority of LADs (73%) had less than 10% of land cover classified as natural. Adult obesity prevalence across included LADs ranged from 13.1% to 29.9%, with a mean of 22.8%. (Table 7.1) The random intercept model showed that 1.7% of the variance in BMI was attributable to between-LAD rather than within-LAD differences.

Table 7.1 Summary of key variables

Individual-level characteristics of sample (n=302,952)	
BMI, mean (SD)	27.5 (4.8)
Number of PA facilities, median (IQR)	2 (0-4)
Distance (m) to nearest fast-food/takeaway store (median, IQR)	996 (560-1726)
Age (years), mean (SD)	56.1 (4.1)
% female	52.6%
% Black Asian or Minority Ethnicity	5.6%
% in paid employment	60.9%
% with household income <£18,000	23.9%
% educated to College or University degree level	33.4%
Area deprivation (Townsend score) , mean (SD)	-1.2 (3.0)
Residential density, median residential addresses per km2 (IQR)	2197 (1393 - 3388)
Local Authority District attributes (n=122)	
'Natural' land cover as % of LAD, median (IQR)	4.9 (0.9 – 10.5)
Obesity prevalence, mean (SD)	22.8% (SD=3.9%)
Gross Disposable Household Income per capita (£ annual), median (IQR)	14981 (12393 – 17862)

3.2 Geographical heterogeneity

Example 1: Neighbourhood PA environment and BMI

Averaged across the LAD-specific models, the mean difference in BMI for each additional PA facility within a one-kilometre street-network distance of participants' homes was

-0.05 kg/m², but the magnitude of the association between number of neighbourhood PA facilities and BMI varied across England (Figure 7.1, & Supplementary Figure 10, Appendix Four). In 92 of the 122 local authority districts, the estimated association was in the expected negative direction. This association was statistically significant at the (arbitrary) 5% threshold in 32 areas, although in several other areas the 95% CI only failed to exclude zero by a small margin. Upon visual inspection, no regional patterning was apparent (Supplementary Figure 10, Appendix Four). For example, areas where the mean BMI difference associated with each additional PA facility near a person's home was at least one standard deviation (0.08) more than average (i.e. a difference of at least 0.13kg/m²) were distributed across the South West, South East, Greater Manchester and the Midlands.

Example 2: Fast-food proximity and BMI

Across LADs, the mean difference in BMI for a 10-fold increase in distance to the nearest fast-food/takeaway store was -0.24 kg/m², and here too the magnitude of the association varied across England (Figure 7.2). The direction of the estimated association was in the expected negative direction in two-thirds of all areas (n=77), however only in 12 districts did the 95% CI around the point estimate exclude zero. There was no obvious regional patterning (Supplementary Figure 11, Appendix Four).

3.3 Effect modification by attributes of the macro environment

Example 1: 'Natural' land cover as a potential modifier of the association between neighbourhood PA environment and BMI

There was some evidence that percentage of land cover classified as 'natural' in a LAD weakly modifies the association between neighbourhood availability of formal PA facilities and BMI. Models testing this cross-level interaction showed the primary association to be stronger among people living in areas with the lowest proportion of natural landcover, for whom each additional PA facility close to home is associated with 0.054 kg/m² lower BMI (95% CI:-0.070, -0.038; P<0.001) compared with a mean BMI difference of -0.032 kg/m² per additional PA facility in those areas with the most natural land cover (95% CI:-0.051, -0.012; P=0.001) ($P_{interaction}$=0.087; Figure 7.3). The fanning out of the lines in Figure 7.3 shows the strengthening association as percentage natural land cover decreases.

Example 2: Obesity prevalence as a potential modifier of the association between fast-food proximity and BMI

There was far weaker evidence that local descriptive obesity norms modify the association between fast-food proximity and BMI. Models testing the cross-level interaction between fast-food proximity and local obesity prevalence estimated the primary association to be slightly stronger among people living in areas with the highest prevalence of adult obesity, for whom a 10-fold increase in the distance to a fast-food store was associated with 0.29 kg/m² lower BMI (95% CI -0.42, -0.17; P<0.001), compared with a mean difference in BMI of -0.21 kg/m² for those in areas where adult obesity was least prevalent (95% CI -0.31, -0.10; P<0.001) ($P_{interaction}$ =0.261; Figure 7.4). As can be seen in Figure 7.4, the gradient for the main association becomes only slightly flatter as obesity prevalence decreases.

Sensitivity analysis

Excluding the 20 Local Authorities represented in UK Biobank that are located in London, we found stronger evidence of a cross-level interaction between neighbourhood availability of PA facilities and the percentage of land cover classified as 'natural' in a LAD

($P_{interaction}$=0.044). Among people living in the areas with the lowest proportion of natural landcover, each additional PA facility close to home was associated with 0.063 kg/m² lower BMI (95% CI:-0.084, -0.042, P<0.001), which was twice the magnitude of the association among people in areas with the most natural land cover (mean BMI difference = -0.031; 95% CI:-0.052, -0.009, P=0.006).

Concerning the association between fast-food proximity and BMI, there was no evidence outside London that adult obesity prevalence in the local authority acted as an effect modifier ($P_{interaction}$=0.730). Across all levels of obesity prevalence, mean BMI was between 0.23 and 0.26 kg/m² lower with each 10-fold increase in distance to nearest fast-food store. Any effect modification by obesity norms in this population appears to be localised to London.

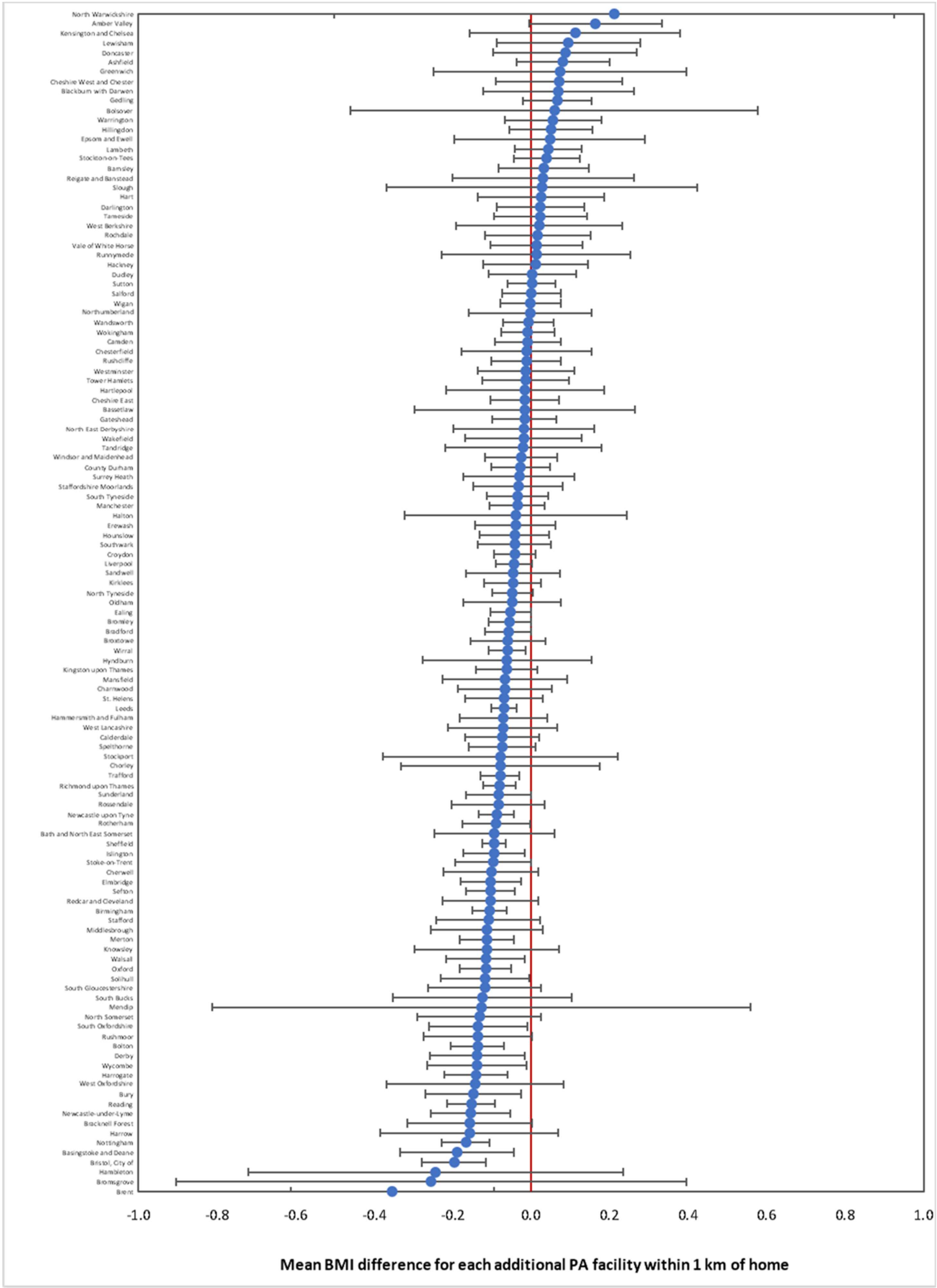


Figure 7.1 Local Authority-specific estimates of mean BMI difference associated with each additional PA facility near home

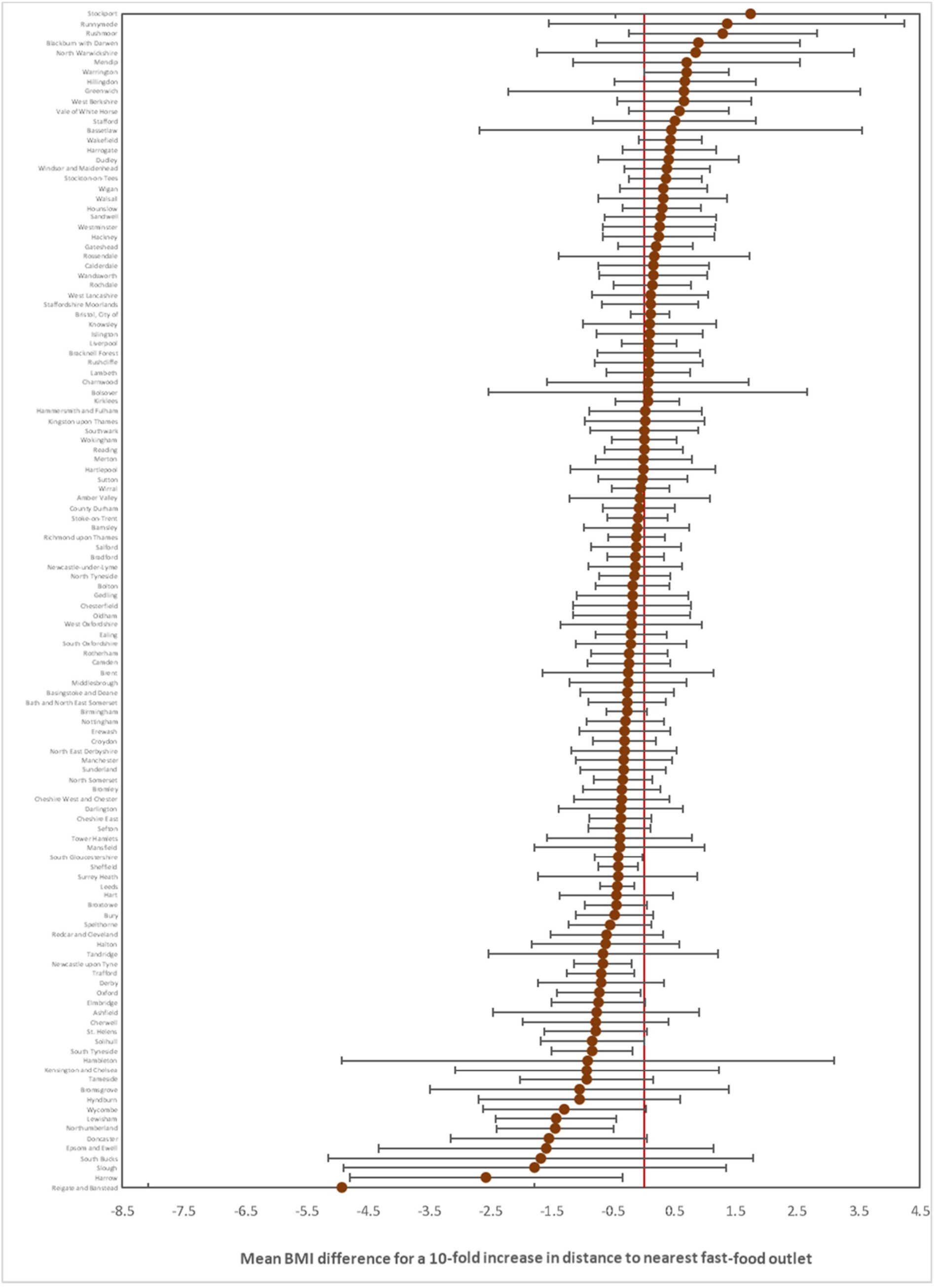


Figure 7.2 Local Authority-specific estimates of mean BMI difference associated with proximity to a fast-food outlet

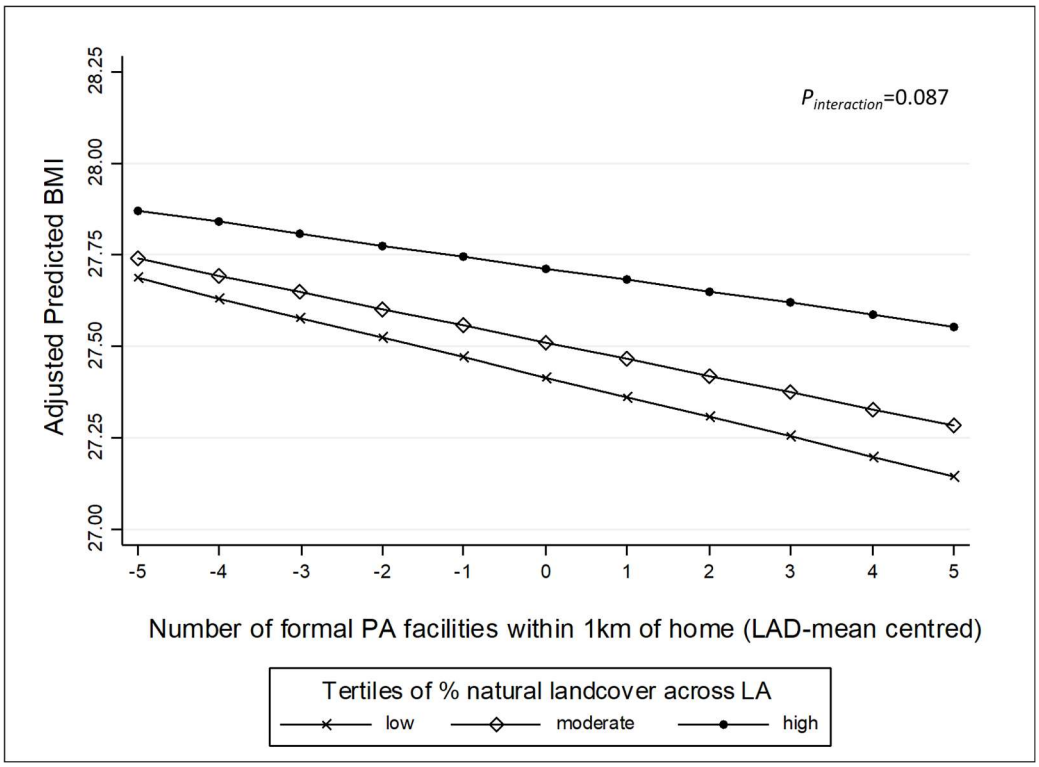


Figure 7.3 Association between neighbourhood availability of PA facilities and BMI, by tertile of percentage 'natural' land cover in local authority

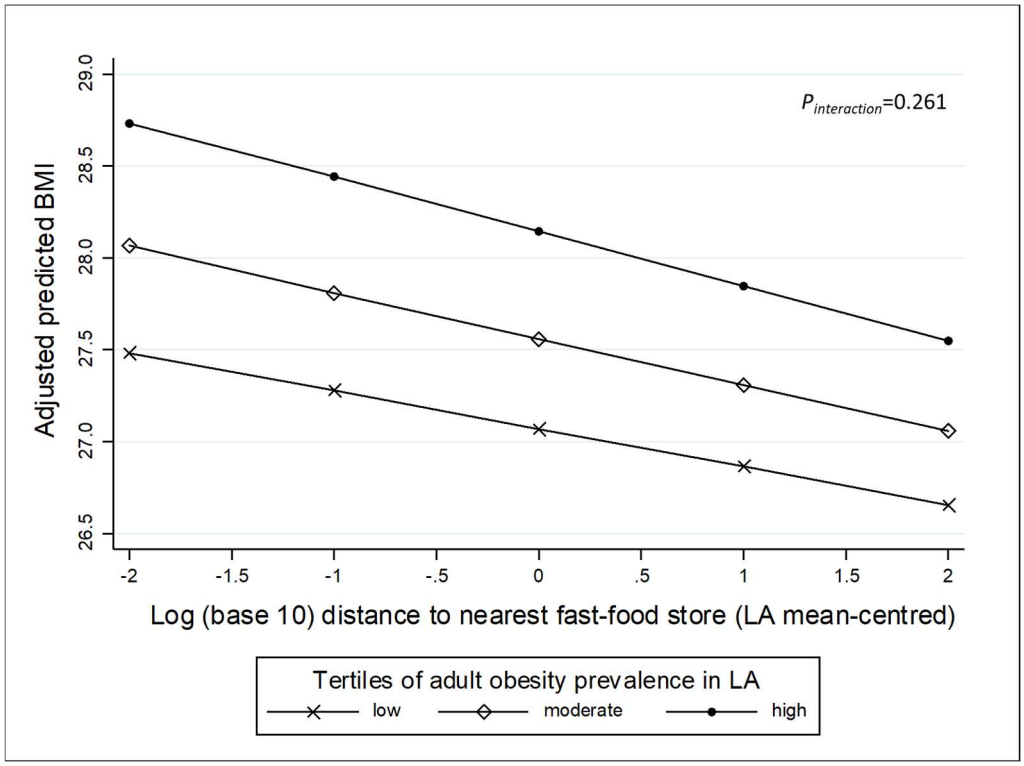


Figure 7.4 Association between neighbourhood fast-food proximity and BMI, by tertile of adult obesity prevalence in local authority

DISCUSSION

In this paper we have demonstrated that associations between the built environment and obesity risk may not be uniform across geographical space, using the examples of neighbourhood availability of PA facilities and neighbourhood fast-food proximity and BMI. Furthermore, we explored the possibility that geographical heterogeneity in these associations may be driven by attributes of the wider contexts in which people live, and found some evidence that the extent to which a given neighbourhood characteristic matters for the health of its residents may depend on features of the larger administrative area in which a neighbourhood is located.

4.1 Interpretation of results

We found that relationships between availability of neighbourhood PA facilities and BMI, and fast-food proximity and BMI, varied from place to place across urban England. While across the sample as a whole we have previously observed a clear, graded, negative association between the availability of PA facilities and adiposity[8], stratification by local authority district reveals that this exists only in a subset of areas, and is of greater magnitude in some areas than others. In some LADs there is no evidence of any such association, while in many others the estimated effect was in the expected direction and the 95% CI only just included the null. Similarly, a strong positive association between fast-food proximity and BMI appears to be present in some areas but not others.

For both relationships, no obvious regional patterning was apparent from visual inspection of geographical heterogeneity; rather, areas with a particularly strong association were dispersed across the country. If the associations had been consistent across space, then we could infer that results of other studies of these relationship are likely to be broadly generalisable from one setting to another, at least within England. However, if as our findings suggest, the associations are geographically heterogeneous, then the generalisability of findings from studies with narrow geographical coverage is undermined, and we should infer that the potential for built environment interventions to be effective might not be universal without careful consideration of context.

Given that we observed the primary associations to vary by a higher-level geographical unit in which neighbourhoods are nested (LADs), we explored the possibility that this may be driven by variation in attributes of those larger areas. In each of our two worked examples, we tested an interaction with a plausible, place-based modifier of the main exposure effect, for which data were publicly available. We observed that a measure of the

physical landscape showed some evidence of modifying the individual-level association between the formal neighbourhood PA environment and BMI, after accounting for covariates at the individual- and area-level. The estimated magnitude of the association was somewhat weaker among people living in areas with more natural landcover. This aligns with our hypothesis that because greater land coverage with natural landscape types provides more opportunities and alternative spaces for PA and may contribute to social norms around PA, less natural landcover may result in greater reliance on and normalisation of the use of formal PA facilities.

With respect to the fast-food environment and BMI, we found little evidence of effect modification by LAD-level adult obesity prevalence – a measure of local descriptive norms. Very few studies have examined the role of local descriptive norms (spatially-defined local prevalence of a trait or behaviour) rather than subjective norms (behaviours or traits of social networks) on health outcomes and behaviours, but those studies we are aware of suggest they may be important[32,50], and they have also been shown to be influential in other domains (e.g. pro-environmental behaviour). We hypothesised that where obesity is less 'normalised', social pressure to maintain a healthy weight might be greater and act to suppress the influence of an unhealthy food environment. Our findings here do not provide strong support for our hypothesis that where obesity prevalence is lower, the association between the fast-food environment and BMI would be weaker, although sensitivity analyses did suggest a different relationship might exist in London areas compared with other areas. Further work to test and isolate any mechanisms such as these is therefore warranted. The influence of local descriptive norms might also be weaker than the influence of subjective norms (e.g. via actual social networks), which need not be constrained by administrative boundaries (or indeed by geography at all)[40], and which we could not examine in this study.

We note that our results indicated that only a small percentage of the total variance in BMI was attributable to differences at the LAD level. Whilst not surprising (most of the variation in BMI would be expected to be explained by individual-level factors, including in this study egocentric neighbourhood characteristics), this does remind us that LAD-level factors are only likely to be one small part of a larger system of determinants.

4.2 Strengths and limitations

Despite the well-known inconsistency of research findings from various settings, and increasing calls for the examination of effect modification in studies of the built environment and health[35,36], this is one of few studies we are aware of that has examined whether and how neighbourhood-obesity associations vary geographically and according to explicitly contextual variables. To our knowledge, this is the first paper to examine geographical heterogeneity across the UK using a single study population and empirically examine possible drivers of that heterogeneity. The paper serves as an exploratory demonstration of the possible presence and drivers of geographical heterogeneity within relationships between built environments and health. Feasibility constraints on large-scale studies of the built environment are likely one reason this phenomenon has rarely been closely examined; UK Biobank and the UKBUMP provided the opportunity to work with a sample sufficiently large to draw comparisons between numerous administrative areas, and link publicly available data for those areas to examine possible effect modification relationships in a way not done before.

This analysis does, however, have numerous limitations, and the results should therefore be viewed principally as a demonstration that geographical heterogeneity in these associations is, in general, a phenomenon requiring closer attention. We provide two examples of the type of investigations that may prove fruitful for understanding drivers of any such heterogeneity. Limitations of this particular study and the data used include possible temporal mismatch between the various data sources used. The UK Biobank baseline assessment period was 2006-2010, and while we matched the external datasets as closely as was possible, and while physical features such as land cover are unlikely to change substantially over just a few years, the various data sources used in this analysis do nonetheless span the period 2005 to 2012. For the purposes of identifying the LAD in which each individual resides, we had to rely on approximate address coordinates, which may have led to some individuals residing near the boundary of a LAD being incorrectly assigned to a neighbouring LAD. While this may introduce some misclassification bias, it is likely that neighbouring LADs are similar to one another in terms of natural landcover and obesity prevalence. Additionally, people living on the edges of LADs may be influenced by characteristics of neighbouring LADs, yet we ignore these potential 'edge effects'[51].

We used pre-classified secondary data to assess the landscape context of each local authority, and were therefore constrained by the classification scheme used there. In the Land Cover Atlas of the UK, urban parks and sport and leisure facilities are jointly

classified as 'urban green', meaning the 'natural' landcover category excludes urban parks. There is an inverse correlation between proportion of urban green and proportion of natural land cover in the data, so if urban parks at the LAD level have a similar modifying effect on the influence of neighbourhood formal PA facilities, we may have underestimated the interaction with 'natural' land cover by not being able to factor in the influence of urban parks.

Our results may also be affected by the uncertain geographic context problem (UGCoP)[52], such that local authority scale may not be the relevant scale for assessing the particular effect modifiers we considered. Our decision to use Local Authority as the scale at which to assess geographical heterogeneity and examine effect modification was partly one of pragmatism – this is a scale at which relevant data are available. But it is also the scale at which many planning and resourcing decisions are made, and the LAD is therefore a potential locus of intervention. So for example, if a local authority was considering a public health intervention that involved increasing the availability of PA facilities on the basis of evidence from UK-wide observational studies or intervention research conducted in another LAD, our results suggest decision makers may be wise to consider LAD-level characteristics that may differ from national averages or from the setting of key studies where evidence has been generated previously. Whether or not the LAD is an aetiologically relevant scale, it is likely to be a relevant scale for planning and resource allocation decisions. One important caveat to the use of LADs is that there are multiple types of LADs in England; some are single towns or cities, others are subdivisions of large cities and urban conurbations, and others still represent less urbanised parts of larger counties. The responsibilities, governance structures and size of each varies. By restricting our analysis to study participants living in urban postcodes, we may have avoided some of the issues this presents, but further research should examine geographical heterogeneity and contextual effect modification at alternative scales and non-administrative boundaries.

Other possible sources of bias include selection bias in UK Biobank as a whole, arising from a low response fraction, the cross-sectional study design adopted here, and structural confounding due to residential segregation and selective migration. If these sources of bias apply differentially across space, they may have given rise to a spurious appearance of geographical heterogeneity. Finally, we assumed a linear relationship between our exposures and outcome, and while this is valid across the sample as a whole, within local authorities it may take other functional forms.

4.3 Future research directions

This paper examines one potentially important scale at which heterogeneity may exist, and two measures of the wider context that might plausibly moderate neighbourhood-BMI associations, but other scales and other place-based variables should be investigated in future research. Other potential contextual modifiers of associations between neighbourhood environments and obesity might include the presence or absence of city-wide initiatives (e.g. around active commuting or healthy food environments), or dominant cultural meanings of PA and food in a region. Further, building on the important contributions to our understanding of neighbourhoods and health that have come from studies of perceived neighbourhood safety and other similar factors[53], examination of city-wide perceptions of safety, crime etc. may also help explain geographical heterogeneity of associations between the objectively observed neighbourhood built environment and health outcomes, especially in relation to PA. Climatic variation between areas may also be an important modifier of the effect of the PA environment.

4.4 Conclusions

Most studies of neighbourhood effects implicitly assume that such effects are universal. We found the relationship between the neighbourhood PA environment and BMI varies from place to place across urban England. If, even in a relatively small country such as England, neighbourhood effects are genuinely not uniform across geographical space, this may have important implications for the generalisability of studies with a narrow geographical focus. However, the possibility – demonstrated in the second part of this paper – that some attributes of the wider context may be important modifiers of neighbourhood effects on health, opens up research avenues for making sense of this geographical heterogeneity. Seeking a deeper understanding of these complex relationships has the potential to inform effective and cost-effective targeting and tailoring of built environment interventions.

ACKNOWLEDGEMENTS

Copyright for the UK Land Cover Atlas data rests with the European Commission, and was produced by the University of Leicester, The Centre for Landscape and Climate Research and Specto Natura and supported by Defra and the European Environment Agency under Grant Agreement 3541/B2012/Ro-GIO/EEA.55055 with funding by the European Union.

REFERENCES

1 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev* 2011; **12**: e95-e106

2 McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int J Behav Nutr Phys Act* 2011; **8**: 125.

3 Malambo P, Kengne AP, De Villiers A, Lambert E V., Puoane T. Built environment, selected risk factors and major cardiovascular disease outcomes: a systematic review. *PLoS One* 2016; **11**: 1–13.

4 Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 2010; **16**: 175–90.

5 Mackenbach JD, Rutter H, Compernolle S, *et al.* Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. *BMC Public Health* 2014; **14**: 233.

6 Black C, Moon G, Baird J. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health Place* 2014; **27**: 229–42.

7 Cummins S, Macintyre S. Food environments and obesity - Neighbourhood or nation? *Int J Epidemiol* 2006; **35**: 100–4.

8 Mason KE, Pearce N, Cummins S. Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank. *Lancet Public Heal* 2018; **3**: e24–33.

9 Ellaway A, Lamb KE, Ferguson NS, Ogilvie D. Associations between access to recreational physical activity facilities and body mass index in Scottish adults. *BMC Public Health* 2016; **16**: 756.

10 Hobbs M, Green MA, Griffiths C, et al. Access and quality of parks and associations with obesity: a cross-sectional study. *SSM - Popul Heal* 2017; **3**: 722–9.

11 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Associations between the combined physical activity environment, socioeconomic status, and obesity: a cross-sectional study. *Perspect Public Health* 2018; **138**: 169–72.

12 Burgoine T, Sarkar C, Webster CJ, Monsivais P. Examining the interaction of fast-food outlet exposure and income on diet and obesity: Evidence from 51,361 UK Biobank participants. *Int J Behav Nutr Phys Act* 2018; **15**: 1–12.

13 Bodicoat DH, Carter P, Comber A, *et al.* Is the number of fast-food outlets in the neighbourhood related to screen-detected type 2 diabetes mellitus and associated risk factors? *Public Health Nutr* 2015; **18**: 1698–705.

14 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 2014; **348**: g1464.

15 Burgoine T, Forouhi NG, Griffin SJ, Brage S, Wareham NJ, Monsivais P. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study. *Am J Clin Nutr* 2016; **103**: 1540–7.

16 Harrison F, Jones AP, van Sluijs EMF, Cassidy A, Bentham G, Griffin SJ. Environmental correlates of adiposity in 9-10 year old children: considering home and school neighbourhoods and routes to school. *Soc Sci Med* 2011; **72**: 1411–9.

17 Burgoine T, Alvanides S, Lake AA. Assessing the obesogenic environment of North East England. *Heal Place* 2011; **17**: 738–47.

18 Griffiths C, Frearson A, Taylor A, Radley D, Cooke C. A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK. *Int J Behav Nutr Phys Act* 2014; **11**: 138.

19 Green MA, Radley D, Lomax N, Morris MA, Griffiths C. Is adolescent body mass index and waist circumference associated with the food environments surrounding schools and homes? A longitudinal analysis. *BMC Public Health* 2018; **18**: 482.

20 Adachi-Mejia AM, Lee C, Lee C, *et al.* Geographic variation in the relationship between body mass index and the built environment. *Prev Med* 2017; **100**: 33–40.

21 Chen M, Creger T, Howard V, Judd SE, Harrington KF, Fontaine KR. Association of community food environment and obesity among US adults: A geographical information system analysis. *J Epidemiol Community Health* 2019; **73**: 148–55.

22 Chaix B, Merlo J, Evans D, Leal C, Havard S. Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Soc Sci Med* 2009; **69**: 1306–10.

23 Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Prev Med* 1999; **29**: 563–70.

24 Wicker P, Downward P, Lera-López F. Does regional disadvantage affect health-related sport and physical activity level? A multi-level analysis of individual behaviour. *Eur J Sport Sci* 2017; **17**: 1350–9.

25 Van Tuyckom C. Macro-environmental factors associated with leisure-time physical activity: a cross-national analysis of EU countries. *Scand J Public Health* 2011; **39**: 419–26.

26 Kokolakakis T, Lera-Lopez F, Castellanos P. Regional differences in sports participation: the case of local authorities in England. *Int J Sport Financ* 2014; **9**: 149–71.

27 Laverty AA, Thompson H, Cetateanu A, Filippidis FT. Macro-environmental factors and physical activity in 28 European Union countries. *Eur J Public Health* 2018; **28**: 300–2.

28 Kokolakakis T, Castellanos-García P, Lera-López F. Differences in formal and informal sports participation at regional level in England. *Int J Sport Policy* 2017; **9**: 491–504.

29 Haase A, Steptoe A, Sallis JF, Wardle J. Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. *Prev Med* 2004; **39**: 182–90.

30 Coutts C, Chapin T, Horner M, Taylor C. County-level effects of green space access on physical activity. *J Phys Act Heal* 2013; **10**: 232–40.

31 Carroll SJ, Niyonsenga T, Coffee NT, Taylor AW, Daniel M. Associations between local descriptive norms for overweight/obesity and insufficient fruit intake, individual-level diet, and 10-year change in body mass index and glycosylated haemoglobin in an Australian cohort. *Int J Behav Nutr Phys Act* 2018; **15**: 1–16.

32 Blok DJ, de Vlas SJ, van Empelen P, Richardus JH, van Lenthe FJ. Changes in smoking, sports participation and overweight: does neighborhood prevalence matter? *Heal Place* 2013; **23**: 33–8.

33 Diez-Roux A V, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010; **1186**: 125–45.

34 Spence JC, Lee RE. Toward a comprehensive model of physical activity. *Psychol Sport Exerc* 2003; **4**: 7–24.

35 Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev* 2011; **12**: 217–30.

36 Diez Roux A V. Conceptual approaches to the study of health disparities. *Annu Rev Public Health* 2012; **33**: 41–58.

37 Arcaya M, Tucker-Seeley R, Kim R, Schnake-Mahl A, So M, Subramanian SV. Research on neighborhood effects on health in the United States: a systematic review of study characteristics. *Soc Sci Med* 2016; **168**: 16–29.

38 Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity* 2015; **23**: 1331–44.

39 Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. *Int J Behav Nutr Phys Act* 2017; **14**: 156.

40 Fornara F, Carrus G, Passafaro P, Bonnes M. Distinguishing the sources of normative influence on proenvironmental behaviors: the role of local norms in household waste recycling. *Gr Process Intergr Relations* 2011; **14**: 623–35.

41 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007.

42 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP): a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21**: 135–48.

43 Cole B, King S, Ogutu B, Palmer D, Smith G, Balzter H. Corine land cover 2012 for the UK, Jersey and Guernsey. 2015.

44 Rae A. A Land Cover Atlas of the UK. 2017. DOI:10.15131/shef.data.5266495.

45 Scholes S, Pickering K, Rayat P, National Centre for Social Research. Neighbourhood Statistics, LA model-based estimates of healthy lifestyles behaviours 2003-05. 2007. https://digital.nhs.uk/data-and-information/publications/statistical/neighbourhood-statistics-mental-health-adults-accessing-nhs-specialist-mental-health-services/neighbourhood-statistics-la-model-based-estimates-of-healthy-lifestyles-behaviours-2003-05 (accessed July 26, 2018).

46 Office for National Statistics. Regional gross disposable household income, UK 1997 to 2016. 2018.

47 Coldwel DF, Evans KL. Contrasting effects of visiting urban greenspace and the countryside on biodiversity knowledge and conservation support. *PLoS One* 2017; **12**: e0174376.

48 Dennis M, James P. Evaluating the relative influence on population health of domestic gardens and green space along a rural-urban gradient. *Landsc Urban Plan* 2017; **157**: 343–51.

49 Enders CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol Methods* 2007; **12**: 121–38.

50 Carroll SJ, Paquet C, Howard NJ, *et al.* Local descriptive norms for overweight/obesity and physical inactivity, features of the built environment, and 10-year change in glycosylated haemoglobin in an Australian population-based biomedical cohort. *Soc Sci Med* 2016; **166**: 233–43.

51 Van Meter EM, Lawson AB, Colabianchi N, *et al.* An evaluation of edge effects in nutritional accessibility and availability measures: a simulation study. *Int J Health Geogr* 2010; **9**: 1–12.

52 Kwan MP. The uncertain geographic context problem. *Ann Assoc Am Geogr* 2012; **102**: 958–68.

53 Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. *Epidemiol Rev* 2009; **31**: 7–20.

Chapter 8. NEIGHBOURHOOD BUILT ENVIRONMENT CHARACTERISTICS, SOCIOECONOMIC POSITION AND CARDIOVASCULAR DISEASE AND CANCER

8.1. Introduction

In the four preceding empirical chapters I undertook cross-sectional analyses of the UK Biobank baseline data, and I focussed on adiposity because of its role along the pathways to health from plausibly 'obesogenic' neighbourhood characteristics related to diet and physical activity. In this final empirical chapter, I make use of administrative health records linked to UK Biobank to conduct longitudinal analyses (thereby seeking to strengthen causal inference), and I expand my focus to more distal outcomes: incident CVD and cancer over a ten-year follow up period. For these outcomes I also consider that neighbourhood environments may exert an influence through pathways other than diet and PA. On those grounds I additionally examine neighbourhood greenspace as an exposure that may also influence, for example, psychosocial pathways to health. Continuing my examination of effect heterogeneity, here I investigate whether socioeconomic factors modify these associations between neighbourhood environment characteristics and prospectively ascertained outcomes.

8.2. Research Paper 5

Neighbourhood built environments, socioeconomic position, and hospital admissions for cardiovascular disease and cancer: a prospective study using UK Biobank and linked administrative records

Note: Supplementary material for this research paper is included in Appendix Five.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	LSH1510923	**Title**	Ms
First Name(s)	Kate		
Surname/Family Name	Mason		
Thesis Title	Where and for whom does the neighbourhood built environment matter for obesity and health?: Examining sources of effect heterogeneity at multiple scales in the UK adult population		
Primary Supervisor	Neil Pearce		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Circulation
Please list the paper's authors in the intended authorship order:	Kate Mason, Neil Pearce, Steven Cummins

Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the analysis and undertook all data management, statistical analysis, and writing of the manuscript. SC and NP contributed to the interpretation of results and drafting of the final manuscript.

SECTION E

Student Signature	[redacted]
Date	01/07/2019

Supervisor Signature	[redacted]
Date	01/07/2019

Neighbourhood built environments, socioeconomic position, and hospital admissions for cardiovascular disease and cancer: a prospective study using UK Biobank and linked administrative records

ABSTRACT

Background: Neighbourhood environments may influence risk of cardiovascular disease (CVD) and cancer, via diet and physical activity (PA) behaviours. However, if the effects of these neighbourhood risk exposures vary by socioeconomic position, efforts to improve population health by improving neighbourhood built environments may widen health inequalities.

Methods: In the UK Biobank cohort we used linked records of hospital admissions to assess the relative hazard of being admitted to hospital with a primary diagnosis of CVD or cancer, according to three characteristics of the neighbourhood built environment: availability of formal PA facilities, proximity of a takeaway/fast-food store, and neighbourhood greenspace. We then examined whether there is evidence of effect modification of the main associations by household income and area deprivation. We used Cox proportional hazards models, adjusted for likely confounding, and calculated relative excess risks due to interaction (RERI) to assess effect modification on the additive scale. We also examined the combined modifying role of income and deprivation.

Results: Mean follow up time was 6.8 years. There were 13,809 incident CVD admissions and 13,935 incident cancer admissions in the sample. Ignoring effect modification, associations between the neighbourhood exposures and CVD- or cancer-related hospital admissions were weak to null. However, there was good evidence of effect modification by both area deprivation and household income. Greater availability of PA facilities close to home was associated with lower risk of CVD-related admissions in more deprived areas, but only among those with higher household incomes. More greenspace was associated with lower risk of cancer-related hospital admission among people in deprived areas, but was not associated with lower risk of CVD-related admission for any group. Area deprivation and household income both modified the association between fast-food proximity and CVD admissions, such that people in low-income households further from a fast-food store had lower risk, but this association was stronger in more affluent areas. Some differences in these results were observed between women and men. Findings were

largely robust to alternative specification of time at risk, and adjustment for additional covariates for which there was some ambiguity about temporal precedence.

Conclusions: Improving deprived neighbourhoods by increasing the number of formal PA facilities, while also ensuring access to these is free or affordable, and increasing the amount of public and private greenspace and limiting the proximity of fast-food outlets to residential areas, may improve health outcomes in the population. Pathways from greenspace to cancer risk require further elucidation and should not be assumed to operate primarily through physical activity. By examining effect modification by multiple socioeconomic indicators in parallel, greater insight can be gained into the potentially different ways in which different aspects of the socioeconomic conditions of people's lives influence their relationship with the built environment and its effects on their health. Understanding this better may help to avoid generating or perpetuating health inequalities when neighbourhood-based built environment interventions are designed.

BACKGROUND

Inadequate physical activity (PA), poor diet and chronic stress are risk factors for conditions such as heart disease, stroke, and various cancers[1–4]. Features of neighbourhood environments have the potential to either support or hamper healthy diet and PA behaviours, and to mitigate or exacerbate the stresses of daily life. Unequal access to healthy neighbourhood resources may therefore result in differential risk of hospital admission due to cardiovascular disease or cancer.

Over the past 25 years, cross-sectional studies have produced inconsistent evidence linking neighbourhood built environment characteristics to cardiovascular disease outcomes[5,6], adiposity and obesity[7], mental health[8], and health behaviours such as PA and diet[9,10]. Recently, findings from longitudinal studies have also contributed to the evidence base (e.g.[11,12]), and these study designs may help to better elucidate the true causal relationships between neighbourhood environments and these outcomes. Longitudinal studies provide greater certainty about the temporal sequence of exposures and outcomes of interest, eliminating the possibility of reverse causation. Linkage of hospital records and mortality registers to population-based cohorts with geographical data on neighbourhood environments provides opportunities to examine whether environmental characteristics of neighbourhoods are associated with objectively recorded, prospective outcomes, consistent with hypothesised relationships.

Numerous features of the neighbourhood built environment are widely hypothesised to influence obesity and, by extension, risk of health outcomes such as CVD and cancer, via either diet or PA. These include, but are not limited to, the retail food environment (including proximity or density of healthy and unhealthy food stores), accessibility of recreation facilities for physical activity (such as public swimming pools, gyms, sports fields), and green spaces such as public parks and private gardens (which may facilitate recreational PA, or functional PA such as gardening or active travel). Greenspace may also offer additional health benefits via other pathways relating to the regulation of stress hormones, improved immune function through exposure to diverse microorganisms, and reduced exposure to air pollution[13], all of which may influence risk of cardiovascular disease, cancer and other chronic diseases. The evidence base for these neighbourhood effects on health remain inconclusive, and the relative importance of different neighbourhood exposures is unclear. If causal neighbourhood effects on health do exist, they are likely to be small, and part of a broader swathe of environmental, social and structural drivers of health behaviours and outcomes, each contributing incrementally to the complex physical and social environments that constrain our ability to make healthy lifestyle choices and mitigate the stresses of modern life.

An important aspect of understanding these relationships is the possibility that they are not uniform across the population, but that some people in some places are more sensitive to their neighbourhood environment than others. Important effects concentrated in particular population subgroups or particular places may be masked by average, population-wide estimates. Socioeconomic differences may be one source of such effect heterogeneity. Results of some studies suggest differential neighbourhood health effects by individual socioeconomic status[14,15] or neighbourhood deprivation[16,17]. These may arise if there is differential access to particular neighbourhood resources in more deprived areas compared with less deprived areas[18], or if there are differential preferences for particular neighbourhood resources according to individual socioeconomic status, regardless of the physical availability of neighbourhood resources e.g. if low-income households tend to make more use of fast-food/takeaway stores, or if access to gyms and leisure centres is restricted to people from high-income households because of membership fees. In the case of greenspace effects on health, differences may arise according to area-level deprivation (rather than individual/household socioeconomic position) if more deprived areas have poorer quality greenspace. On the other hand, if greenspace promotes health without an attendant increase in financial costs to the individual, then access to greenspace may offset inequitable access to formal PA facilities, and therefore have a larger effect in deprived

areas or for low-income households. Regardless of the direction of any such heterogeneity of effect, it remains a poorly understood aspect of the relationships between neighbourhood environments and health. If differential benefits or harms of neighbourhood characteristics are observed by household income, or other measures of individual socioeconomic status, or by neighbourhood deprivation, then any efforts to improve population health by improving neighbourhood built environments (e.g. increasing availability of PA facilities or reducing the number of fast-food outlets near residential areas) may widen health inequalities if they are blind to socially differential impacts[19].

In this paper we use baseline UK Biobank data on neighbourhood exposures to PA facilities, greenspace and fast-food stores, linked to records of hospital admissions up to January 2016, to examine (1) the relative hazard of being admitted to hospital with a primary diagnosis of cardiovascular disease or cancer, according to exposure to each of the neighbourhood characteristics, and (2) whether there is evidence of effect modification of those associations by household income and/or area deprivation.

METHODS

Study design and data collection

We used data from UK Biobank (project 17380), the scientific rationale, study design and survey methods for which have been described elsewhere[20]. More than half a million individuals were recruited to visit one of 22 UK Biobank assessment centres across the UK between 2006 and 2010. All individuals aged 40–69 years living within a 25-mile radius of an assessment centre and listed on National Health Service (NHS) patient registers were invited to participate in the study. The age range was chosen by UK Biobank as an important period for the development of many chronic diseases.

Neighbourhood environment data

Linked to UK Biobank is a high-resolution spatial database of a range of objectively measured characteristics of the physical environment surrounding each participant's exact residential address, known as the UK Biobank Urban Morphometric Platform (UKBUMP). Environmental data in UKBUMP were derived, using automated processes, from multiple pre-existing sources roughly contemporaneous with the individual baseline assessment[21]. Over time, as researchers work with UK Biobank, new linked data are being made available

to the research community, including additional environmental measures of greenspace[22] that we have used here in addition to measures from the original UKBUMP.

Linked hospital admissions data

Ongoing prospective linkage of the cohort to administrative health records is a key feature of the UK Biobank resource. At the time of analysis, linked Hospital Episode Statistics were available up to January 2016. These contain information on hospital admissions coded using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). We used these data to identify incident admissions to hospital for cardiovascular disease and for cancer.

Outcomes

Outcomes were any hospital admission for which the primary diagnosis is recorded as cardiovascular disease (ICD-10 codes I10-I25, I46,I48,I50,I60-79) or cancer (ICD-10 codes C00-C97). CVD and cancer were examined separately. In a set of secondary analyses we examine breast cancer and colorectal cancers specifically, as these have some of the strongest links to physical activity[23] and, to a lesser extent, diet[24], two of the most likely mediators of the neighbourhood effects being examined.

Neighbourhood exposures

Three measures of the neighbourhood built environment were examined. To account for skewed distributions and to facilitate a categorical approach to the analysis of effect modification[25], each exposure was split into four categories. The exposures we examined were:

(1) Availability of PA facilities: number of formal PA facilities within a one-kilometre street-network distance of each participant's home address, categorised as 0, 1, 2-3, or 4 or more.

(2) Fast-food proximity: street-network distance in metres from participants' home address to the nearest 'hot/cold fast-food outlet/takeaway', categorised as <500 m, 500-999 m, 1000-1999 m, 2000 m+.

(3) Greenspace: percentage of 300 m Euclidean buffer around home address classified as 'greenspace' or 'domestic garden' in the Generalised Land Use Database. Combining 'greenspace' and 'gardens' is consistent with previous research using the GLUD to examine relationships with health[11]. A 300 m buffer was chosen to capture greenspace in the

immediate vicinity of a person's home. There is some evidence that 300 m is a distance from home beyond which the use of green spaces quickly declines[26,27], and it has been proposed in the UK as a benchmark for greenspace provision[28]. Greenspace was grouped into quartiles.

Exposures (1) and (2) were derived in the UKBUMP from OS AddressBase Premium 2012[21], while (3) was derived by Wheeler et al from the Generalised Land Use Database 2005[22]. We restricted the analyses to people residing in England, because the greenspace data for exposure (3) were not available for UK Biobank participants in Wales and Scotland.

Potential effect modifiers

We examined whether the association between each neighbourhood exposure and each outcome was modified by binary indicators for annual, pre-tax household income (<£31,000 or ≥£31,000) and area deprivation (most deprived 40% of UK census output areas vs least deprived 60%, based on the Townsend index). When testing for effect modification, household income and area deprivation were combined with each primary exposure into a categorical variable capturing all combinations of levels of the exposure and potential modifier, with a single reference category (see below for details). We also examined the combined modifying role of income and deprivation. Area-based and individual indicators of socioeconomic disadvantage have been shown to contribute to health outcomes independently of one another, providing a rationale for examining them both in parallel and in combination[29].

Potential confounders

We identified potential confounders of the primary associations as age (years), sex (binary), ethnicity (White/non-White), educational qualifications (College or University degree/A levels/AS levels or equivalent/O levels or below/other), employment status (paid work, retired, unable to work, unemployed, or other), urban/non-urban status, UK Biobank assessment centre, and neighbourhood residential density (count of residential dwellings within a 1-km street-network buffer of home address, log transformed). Annual household income (<£18,000, £18,000–30,999, £31,000–51,999, at least £52,000) and area deprivation (Townsend score) were also included as possible confounders in any models where they were not being tested as a potential effect modifier. We also adjusted models for smoking status (current/previous/never), alcohol intake frequency (less than/at least 3 times per week) as these are important risk factors for the outcomes and may be correlated with neighbourhood, and number of years living at current (baseline) address

to at least partially condition on pre-baseline exposure to neighbourhood environment, which could act as a confounder.

Statistical analysis

Of the 502,617 participants in UK Biobank for whom some data were available, 502,544 remained after withdrawals were excluded, and 355,691 of these individuals lived in England and had complete data on covariates and data for at least one measure of the neighbourhood environment. Of these, we excluded 19,535 individuals who had reported prior cardiovascular events from analyses involving CVD outcomes, and 29,112 individuals with a previous cancer diagnosis at baseline from analyses involving cancer outcomes. This left a possible N= 336,156 for the analyses of CVD outcomes and N=326,579 for the analyses of cancer outcomes (Figure 8.1). The final analytic sample sizes varied according to availability of the neighbourhood variable under examination. The maximum follow-up time after baseline assessment was 9.8 years, but varied according to the date of an individual's recruitment to the study.

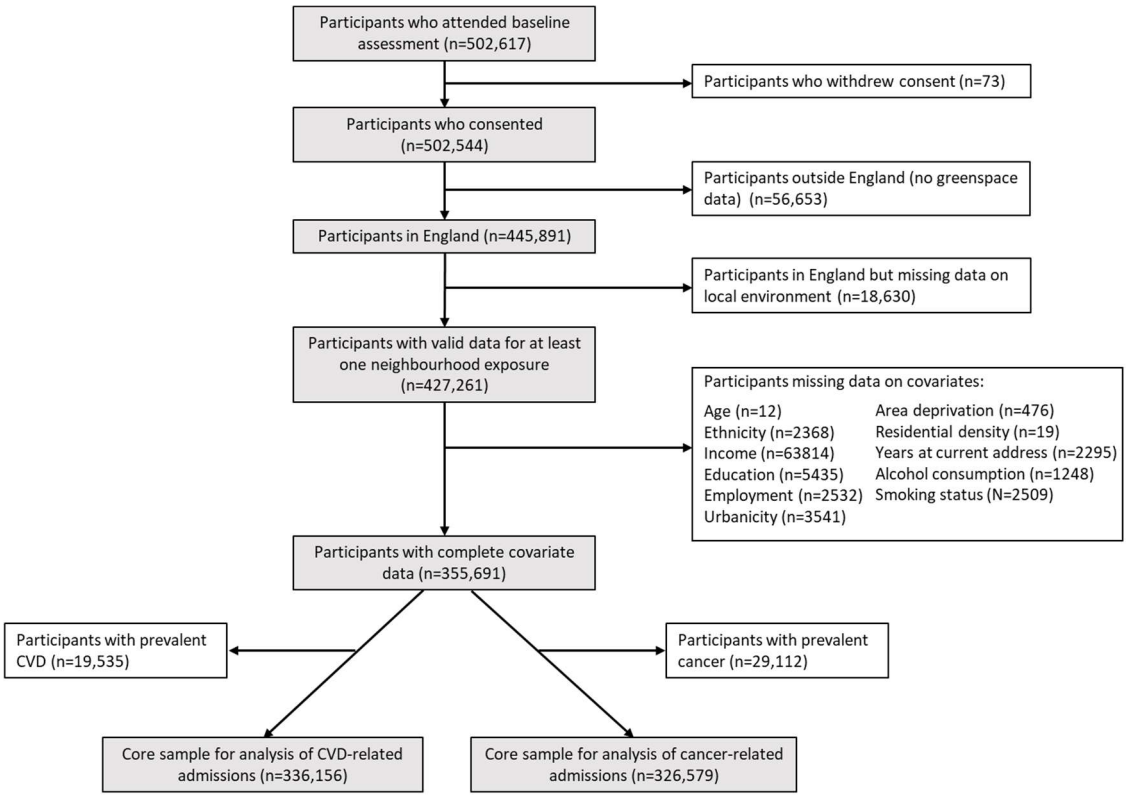


Figure 8.1 UK Biobank sample for analyses

Baseline characteristics were summarised by the mean (and standard deviation) or median (and interquartile range) for continuous variables and number (and percent) for categorical variables. We then examined associations between neighbourhood exposure and incident hospital admission due to CVD or cancer following baseline assessment using multivariable Cox proportional hazard models, with adjustment for potential confounders and censoring for death. Results are expressed as hazard ratios (HRs) and 95% confidence intervals (95% CI). The proportional hazards assumption was tested by visual inspection of adjusted log-log plots (Supplementary Figure 15, Appendix Five). The reference categories for each neighbourhood exposure are the hypothetically least health-promoting (lowest availability of PA facilities, shortest distance to nearest fast-food store, least greenspace).

We examined whether the primary associations were modified by area deprivation and household income. In line with STROBE recommendations[30] and using the method described by Li and Chambless[31] and VanderWeele[25], the relative excess risk due to interaction (RERI) was calculated to assess effect modification on the additive scale. When dealing with binary and time-to-event outcomes, the decision to examine effect modification on either the multiplicative or the additive scale has implications for interpretation. The additive scale provides important information about the potential public health consequence of intervening on the exposure, for different strata of the effect modifier. This is not information we can glean directly from an examination of effect modification on the multiplicative scale, because measures of effect modification on the multiplicative scale ignore potentially different baseline risks within strata of the effect modifier[25]. The RERI is calculated by estimating the HR for each combination of the exposure and potential modifier values relative to a single reference category, in this case the least hypothetically health-promoting level of the respective neighbourhood variable (no PA facilities; <500 m from nearest fast-food store; or quartile with least greenspace), and either low income (<£31,000) or more deprived area (home address located in a census output area in the most deprived 40% of all UK areas). In other words, the reference category in each analysis is the group expected to have the highest baseline risk of the outcome. From this model, the RERI is calculated as:

$$RERI = HR_{11} - HR_{10} - HR_{01} + 1$$

For the model assessing effect modification of PA facility availability by household income, for example, HR_{11} represents the HR (relative to the reference category) for people in high-income households (at least £31,000 per year) and who have the highest level of

neighbourhood availability of PA facilities (4 or more within 1000 m of home); HR_{10} represents the HR for people in low-income households with 4 or more PA facilities near home; and HR_{01} is the HR for people in high-income households with no PA facilities near home.

For the models of the other neighbourhood exposures, and models of effect modification by area deprivation, subscript *1* represents those most exposed to the potentially health-promoting neighbourhood exposure and less deprived areas, respectively. As such, a RERI value greater than zero – which implies a positive departure from additivity – suggests that in this case any estimated protective effect of the neighbourhood variable among people in low-income households or in more deprived areas is greater than the estimated protective effect among people from high-income households or less deprived areas. In contrast, a RERI<0 suggests any protective effect of the neighbourhood variable is greater in the high-income/less deprived group.

All analyses were conducted using Stata v14.2 (StataCorp LP, College Station, TX, USA).

Sensitivity analyses

The spatial data used in the creation of the UKBUMP to ascertain the neighbourhood food and physical activity exposures were recorded in 2012, just after the baseline data collection period[21], and while it is assumed that neighbourhood exposure will be sufficiently constant over this period as to not unduly influence the results, we check this assumption by conducting a sensitivity analysis in which follow-up is restricted to the period from 2012 onwards for all participants, rather than from the baseline assessment date (which could be as early as 2006).

The primary analyses were not adjusted for baseline hypertension or BMI, or medications for hypertension or cholesterol, because of ambiguity regarding temporal precedence. Although these are important risk factors for CVD especially, rather than being confounders they may be on the causal pathways from neighbourhood environment to CVD or cancer if neighbourhood exposure predates them, which it is likely to in this largely residentially stable population (median time living at baseline address was 16 years). We instead conduct sensitivity analyses to examine whether adjusting for these variables influences the primary findings. We also report the main associations stratified by sex.

RESULTS

Descriptive

Table 8.1 summarises the characteristics of the study participants at the baseline assessment. The sample has a mean age of 56 years at baseline and was predominantly of White ethnicity and urban dwelling. Reflecting the age of the sample, just over half were educated to no higher than O levels, and six in every ten were employed at baseline. Participants were evenly distributed across income categories, with roughly half living in households with an annual gross income below £31,000, while 29% lived in the more deprived 40% of areas in the UK.

The mean follow-up time for participants was 6.8 years. Over the follow-up period, 13,809 individuals (4.11%) were admitted to hospital with CVD, and 13,935 (4.27%) were admitted with cancer (Table 8.2). Proportionally, there were more hospital admissions for both outcomes among people from low-income households, whereas admissions were similar across levels of area deprivation.

Table 8.1 Sample characteristics

	CVD analysis	*Cancer analysis*
Total N	336,156	326,579
% female	53.94%	51.65%
Age (years) (mean, SD)	55.94 (8.07)	55.96 (8.09)
Ethnicity (% non-White)	4.94%	5.16%
% Urban	85.48%	85.67%
Education (%)		
College or University degree	34.95%	34.34%
A levels/AS levels or equivalent	11.80%	11.65%
O levels or below/other	53.26%	54.01%
Employment status		
Paid work	62.06%	61.76%
Retired	30.29%	30.31%
Unable to work	2.44%	2.74%
Unemployed	1.53%	1.57%
Other	3.68%	3.62%
Residential density (residential sites per 1000m buffer) (median, IQR)	1915 (1102–3129)	1921 (1109–3130)
Years at current address (median, IQR)	15 (7–25)	15 (7–25)
Area deprivation (mean Townsend score)	-1.46 (2.95)	-1.42 (2.97)
Area deprivation (% in two most deprived quintiles of the UK)	28.86%	29.34%
Household income		
<£18,000	21.48%	22.18%
£18 000–30 999	25.46%	25.37%
£31 000–51 999	26.56%	26.29%
£52 000 or more	26.50%	26.17%
Smoking status		
Current	10.13%	10.35%
Previous	34.42%	34.82%
Never	55.45%	54.83%
Frequency of alcohol consumption (% ≥3 times per week)	45.44%	45.27%

Table 8.2 Hospital admissions by household income and area deprivation

		CVD		*Cancer*
	N	Admissions (%)	N	Admissions (%)
Total	336,156	13,809 (4.11)	326,579	13,935 (4.27)
Household income (annual pre-tax)				
<£31,000	157,818	8185 (8.19)	155,283	8126 (5.23)
£31,000 or more	178,338	5624 (3.16)	171,296	5809 (3.39)
Area deprivation				
More deprived	97,005	4183 (4.31)	95,818	4043 (4.22)
Less deprived	239,151	9626 (4.03)	230,761	9892 (4.29)

* Self-reported average total household income before tax

** 'More deprived' refers to people living in areas in the top two most deprived quintiles of the UK, based on the Townsend index.

The remainder of this results section is split into two parts: first are the results of the analyses for the three exposures and CVD-related hospital admissions, followed by the results of the analyses for the three exposures and cancer-related hospital admissions.

Cardiovascular disease admissions

a) Associations between neighbourhood characteristics and CVD admissions

Figure 8.2 summarises the hazard ratios for hospital admissions due to CVD associated with each of the three neighbourhood environment measures, across the sample as a whole. While 95% CIs for all associations included the null value of 1.0, there was some indication of a weak trend toward decreasing hazard of CVD-related hospital admission with increasing distance to the nearest takeaway/fast-food store. With each category decrease in proximity, the HR for hospital admission moved further away from 1.0, and those living further from a fast food store had a 4% reduced hazard compared with those living closest (HR=0.96; 95%CI: 0.91-1.02). For neighbourhood availability of PA facilities and greenspace, there was little to no evidence of an association with risk of CVD-related admission when averaging across the study population as a whole.

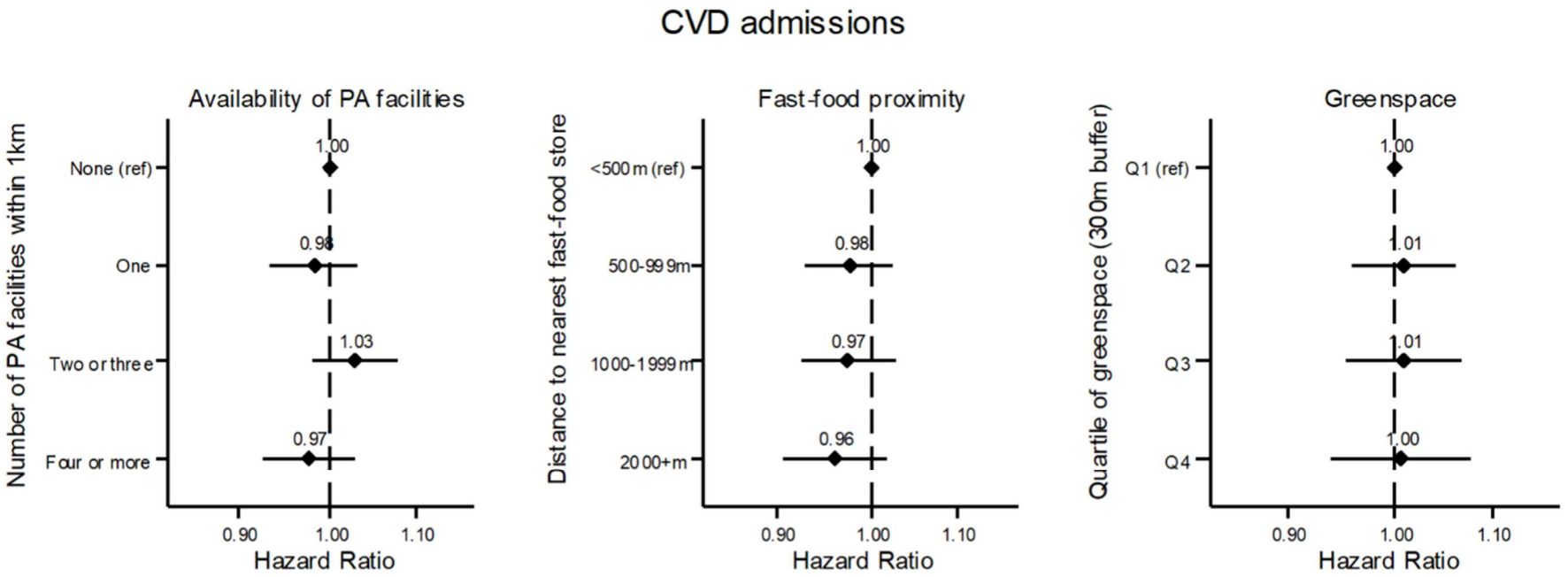


Figure 8.2 Adjusted hazard ratios for hospital admission due to CVD, by availability of formal PA facilities, proximity to nearest fast-food/takeaway store, and neighbourhood greenspace

Note: Models are adjusted for age, sex, ethnicity, education, household income, employment status, urban/non-urban, assessment area, residential density, smoking status, alcohol intake, and number of years living at home address. Plots from sex-stratified models can be found in supplementary material (Appendix Five)

b) Modification of the associations between neighbourhood characteristics and CVD-related hospital admissions, by income and area deprivation

Table 8.3 shows results of the analyses of effect modification by household income and area deprivation, for the associations between each neighbourhood exposure and CVD-related hospital admissions. For availability of PA facilities within a kilometre of home, there was some evidence of effect modification by area deprivation on the additive scale, with a RERI of 0.088 indicating a departure from additivity and a stronger association among people from more deprived areas. Among those living in areas in the two quintiles of greatest deprivation, the stratum-specific hazard ratio for those with at least four PA facilities near home was 0.90 (95%CI: 0.82–0.99), while we observed no association among people in less deprived areas (HR=1.01; 95%CI: 0.95–1.07), suggesting a greater protective effect in more deprived areas. The RERI for effect modification by household income, whilst not as large (RERI=-0.077), indicated a negative departure from additivity, suggesting effect modification in the other direction and a stronger association between PA facilities and CVD among those from higher income households. In stratified models the hazard ratio for those in higher income households with at least four PA facilities near home compared with none was 0.93 (95%CI: 0.85–1.00), while no such association was observed among people in lower income households (HR=1.00; 95%CI: 0.94–1.07).

For fast-food proximity, we observed effect modification by both household income (RERI=0.076) and area deprivation (RERI=-0.104), but in the opposite direction from what was observed for PA facilities (Table 8.3). Reduced access to fast-food/takeaway stores might have the biggest impact for low-income households rather than higher income households, and mostly in less deprived areas. In stratified models, for people in low-income households, living at least 2 km from a fast-food store was associated with a 7% reduction in the hazard of CVD-related hospital admission compared with living within 500 m of a fast-food store (HR-0.93; 95% CI: 0.86-1.00), while no such association was observed among higher income households (HR=1.01; 95% CI: 0.92-1.11). But for people living in more deprived areas, there was no observed protection afforded by living further from a fast-food store, whereas for those in more affluent areas, living at least 2km from a store was associated with an 8% reduced hazard of hospitalisation for CVD (HR=0.92; 95% CI: 0.85-0.99).

Household income and area deprivation did not appear to modify associations between greenspace and CVD-related hospital admissions. There was some weak evidence of effect

modification by income (RERI=0.070), but the stratum-specific HRs were null in both income groups (Table 8.3).

Combining area deprivation and household income, the hazard ratio for at least four PA facilities (compared with none) is smallest among people from high-income households living in more deprived areas (Table 8.4). Among this group, the hazard of being hospitalised for CVD was 22% lower for people who had at least four PA facilities within a kilometre of their home, compared with having no nearby PA facilities (HR=0.78, 95% CI: 0.65-0.94).

For fast-food proximity, a beneficial association of living further from a fast-food/takeaway store was only observed among low-income households in affluent areas, where the hazard of CVD-related admission was 12% lower among people living ≥2km from a fast-food store than among people living <500 m from one (HR=0.88, 95%CI:0.80-0.97; Table 8.4).

For neighbourhood greenspace, confidence intervals around all HRs were too wide to conclude that there was a CVD-related benefit of more greenspace in any of the income/deprivation combinations (Table 8.4).

c) Sex differences

Examining sex differences in the primary results, we found that for the population as a whole, an association between the formal PA environment and CVD admissions was observed among women but not men (Supplementary Table 8, Appendix Five). No effect modification was observed for women. The lack of association among men, however, obscured socioeconomic differences: men in deprived areas and men in higher income households had reduced hazard of CVD-related hospital admission when they had greater neighbourhood availability of PA facilities.

For fast-food proximity and CVD-related admission, the primary results were reflected in men, but no associations were observed for women, for any socioeconomic group (Supplementary Table 9, Appendix Five). The lack of evidence for any association between neighbourhood greenspace and CVD-related admissions, across all socioeconomic strata, was consistent for men and women (Supplementary Table 10, Appendix Five).

Table 8.3 Modification of the association between built environment variables and hospital admissions due to CVD, by household income and area deprivation

CVD-related admissions	*Annual household income**		*Area deprivation***	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None (ref)	1.00	0.99 (0.92, 1.05) P=0.666	1.00	0.88 (0.81, 0.95) P=0.001
One	1.00 (0.94, 1.07) P=0.956	0.94 (0.87, 1.01) P=0.086	0.93 (0.84, 1.03) P=0.152	0.88 (0.81, 0.95) P=0.002
2-3	1.04 (0.98, 1.10) P=0.214	0.99 (0.92, 1.07) P=0.834	0.98 (0.90, 1.07) P=0.696	0.92 (0.85, 1.00) P=0.040
4 or more	1.00 (0.94, 1.07) P=0.886	0.92 (0.85, 0.98) P=0.015	0.92 (0.84, 1.00) P=0.051	0.88 (0.81, 0.96) P=0.004
Stratum-specific HRs (4+ facilities vs 0)	1.00 (0.94, 1.07) P=0.957	0.93 (0.85, 1.00) P=0.062	0.90 (0.82, 0.99) P=0.038	1.01 (0.95, 1.07) P=0.821
Relative excess risk due to interaction (RERI) (95% CI)	-0.077 (-0.167, 0.013) P=0.092		0.088 (-0.004, 0.181) P=0.061	
Fast-food proximity				
Closer than 500m (ref)	1.00	0.92 (0.84, 1.00) P=0.041	1.00	0.98 (0.91, 1.06) P=0.615
500-999m	0.97 (0.91, 1.03) P=0.353	0.90 (0.84, 0.98) P=0.009	1.01 (0.94, 1.09) P=0.771	0.92 (0.86, 0.99) P=0.026
1000-1999m	0.97 (0.91, 1.03) P=0.308	0.91 (0.84, 0.98) P=0.012	0.97 (0.90, 1.06) P=0.548	0.93 (0.87, 1.00) P=0.055
At least 2000m	0.93 (0.87, 1.00) P=0.048	0.92 (0.85, 1.00) P=0.046	1.03 (0.93, 1.14) P=0.595	0.90 (0.84, 0.98) P=0.011
Stratum-specific HRs (≥2000m vs <500m)	0.93 (0.86, 1.00) P=0.047	1.01 (0.92, 1.11) P=0.854	1.03 (0.93, 1.16) P=0.547	0.92 (0.85, 0.99) P=0.019
Relative excess risk due to interaction (RERI) (95% CI)	0.076 (-0.020, 0.171) P=0.122		-0.104 (-0.225, 0.017) P=0.093	
Greenspace				
Q1 (least greenspace) (ref)	1.00	0.91 (0.84, 0.98) P=0.015	1.00	0.98 (0.91, 1.05) P=0.547
Q2	1.00 (0.94, 1.06) P=0.897	0.94 (0.87, 1.01) P=0.094	1.02 (0.95, 1.10) P=0.560	0.94 (0.88, 1.00) P=0.070
Q3	0.99 (0.93, 1.06) P=0.790	0.95 (0.88, 1.02) P=0.154	1.05 (0.95, 1.15) P=0.340	0.93 (0.87, 1.00) P=0.036
Q4 (most greenspace)	0.98 (0.90, 1.05) P=0.529	0.95 (0.88, 1.04) P=0.275	0.98 (0.87, 1.11) P=0.761	0.94 (0.87, 1.01) P=0.104
Stratum-specific HRs (Q4 vs Q1)	0.99 (0.91, 1.08) P=0.844	1.02 (0.91, 1.13) P=0.747	0.96 (0.83, 1.12) P=0.605	0.96 (0.88, 1.04) P=0.295
Relative excess risk due to interaction (RERI) (95% CI)	0.070 (-0.023, 0.164) P=0.142		-0.019 (-0.154, 0.115) P=0.778	

* Self-reported average total household income before tax. ** 'More deprived' refers to people living in areas in the top two most deprived quintiles of the UK, based on the Townsend index. Q= quartile

Note: RERI>0 indicates positive effect modification and a departure from additivity. Here, because the reference category is people with the hypothetically least-healthy level of neighbourhood exposure and with low household income or living in more deprived areas, a positive departure from additivity suggests any estimated *protective* effect of the neighbourhood variable is *weaker* among people in higher income households or in less deprived areas than it is among people from low income households or more deprived areas, and therefore that the latter stand to gain more from an intervention. In contrast, a RERI<0 suggests any protective effect of the neighbourhood variable is stronger in the high-income/less deprived group.

Table 8.4 Association between neighbourhood characteristics and CVD-related hospital admissions, stratified by household income and area deprivation in combination

CVD-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None (ref)	1.00	1.00	1.00	1.00
One	0.96 (0.85, 1.08) P=0.461	0.84 (0.69, 1.03) P=0.103	1.02 (0.95, 1.10) P=0.574	0.96 (0.88, 1.05) P=0.372
2-3	1.00 (0.90, 1.11) P=0.977	0.90 (0.75, 1.08) P=0.267	1.06 (0.98, 1.14) P=0.144	1.01 (0.93, 1.10) P=0.745
4 or more	0.95 (0.85, 1.07) P=0.398	0.78 (0.65, 0.94) P=0.010	1.03 (0.95, 1.12) P=0.481	0.96 (0.88, 1.05) P=0.390
Fast-food proximity				
Closer than 500m (ref)	1.00	1.00	1.00	1.00
500-999m	1.03 (0.94, 1.13) P=0.547	0.99 (0.85, 1.14) P=0.872	0.93 (0.85, 1.01) P=0.085	0.97 (0.88, 1.08) P=0.613
1000-1999m	0.96 (0.87, 1.07) P=0.481	1.04 (0.88, 1.23) P=0.670	0.95 (0.87, 1.04) P=0.298	0.96 (0.87, 1.07) P=0.454
At least 2000m	1.00 (0.88, 1.14) P=0.969	1.12 (0.90, 1.39) P=0.301	0.88 (0.80, 0.97) P=0.013	0.98 (0.88, 1.09) P=0.724
Greenspace				
Q1 (least greenspace) (ref)	1.00	1.00	1.00	1.00
Q2	1.01 (0.92, 1.11) P=0.798	1.10 (0.95, 1.28) P=0.210	0.99 (0.90, 1.08) P=0.803	0.96 (0.86, 1.07) P=0.467
Q3	1.05 (0.93, 1.17) P=0.428	1.06 (0.86, 1.29) P=0.600	0.97 (0.88, 1.07) P=0.550	0.98 (0.88, 1.10) P=0.728
Q4 (most greenspace)	0.93 (0.78, 1.12) P=0.449	1.08 (0.81, 1.44) P=0.591	0.98 (0.88, 1.10) P=0.786	0.99 (0.87, 1.12) P=0.832

* Self-reported average total household income before tax

** 'More deprived' refers to people living in areas in the top two most deprived quintiles of the UK, based on the Townsend index.

Q = quartile

Cancer admissions

a) Associations between neighbourhood characteristics and cancer admissions

Figure 8.3 summarises the hazard ratios for hospital admissions due to cancer associated with each of the three neighbourhood environment measures, across the sample as a whole. While 95% CIs for all associations included the null value of 1.0, there was some indication of a slightly lower hazard of cancer-related hospital admission among those people with at least four PA facilities within one kilometre of their home, compared to people with no nearby formal PA facilities (HR=0.96; 95%CI: 0.91-1.01), but no evidence that one, two or three facilities offers a benefit over none. For fast-food proximity and neighbourhood greenspace, we observed no association with risk of CVD-related admission when averaging across the study population as a whole.

b) Modification of the associations between neighbourhood characteristics and cancer-related hospital admissions, by income and area deprivation

The association between PA facilities and cancer does not appear to be modified by income or by area deprivation. Stratum-specific HRs were similar across socioeconomic groups, and RERIs were close to zero for both potential effect modifiers (Table 8.5).

In contrast, there was some evidence of effect modification by socioeconomic conditions for the associations between the other neighbourhood exposures and cancer. The most marked evidence was for a modifying effect of area deprivation on the association between greenspace and cancer-related admissions. In that case, the positive departure from additivity indicated by the RERI of 0.170 suggests the public health benefits of increased exposure to neighbourhood greenspace may be greater in more deprived areas (Table 8.5). In more deprived areas, the stratum-specific HRs estimate a 16% lower hazard of cancer-related hospitalisation among those in the most green quartile compared with those from the least green quartile (HR=0.84; 95% CI: 0.71-0.98), while no association was observed among people living in more affluent areas. A similar pattern was observed for fast-food proximity and cancer-related admissions, albeit with a smaller and non-significant departure from additivity (RERI=0.070) and a smaller estimated reduction in hazard among the more deprived areas (HR=0.93; 95% CI: 0.83-1.04). For household income, although the RERIs for both fast-food proximity and greenspace did indicate some departure from additivity, the stratum-specific HRs suggested there was no meaningful association between these neighbourhood exposures and cancer-related admissions in either income group (Table 8.5).

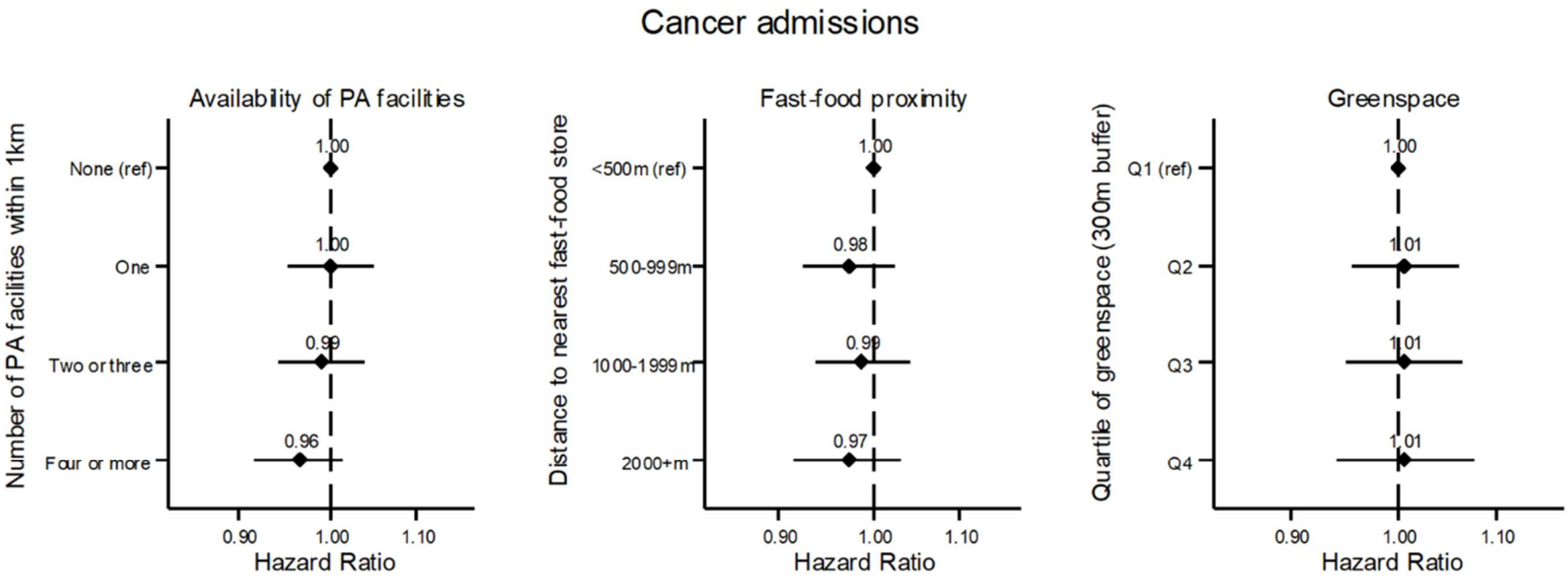


Figure 8.3 Adjusted hazard ratios for hospital admission due to cancer, by availability of formal PA facilities, proximity to nearest fast-food/takeaway store, and neighbourhood greenspace

Note: Models are adjusted for age, sex, ethnicity, education, household income, employment status, urban/non-urban, assessment area, residential density, smoking status, alcohol intake, and number of years living at home address. Plots from sex-stratified models can be found in supplementary material (Appendix Five).

Table 8.5 Modification of the association between built environment variables and hospital admissions due to cancer, by household income and area deprivation

Cancer-related admissions	*Annual household income**		*Area deprivation***	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None (ref)	1.00	0.96 (0.90, 1.02) P=0.209	1.00	1.02 (0.94, 1.11) P=0.605
One	0.98 (0.92, 1.05) P=0.588	0.98 (0.91, 1.06) P=0.608	1.07 (0.97, 1.19) P=0.175	1.00 (0.92, 1.09) P=0.961
2-3	1.00 (0.94, 1.06) P=0.918	0.94 (0.87, 1.00) P=0.066	1.06 (0.96, 1.16) P=0.249	0.99 (0.91, 1.08) P=0.844
4 or more	0.99 (0.93, 1.05) P=0.676	0.89 (0.83, 0.96) P=0.001	1.01 (0.92, 1.11) P=0.846	0.98 (0.90, 1.06) P=0.581
Stratum-specific HRs (4+ facilities vs 0)	0.97 (0.91, 1.04) P=0.395	0.95 (0.87, 1.03) P=0.184	1.02 (0.92, 1.13) P=0.673	0.95 (0.89, 1.01) P=0.110
Relative excess risk due to interaction (RERI)	-0.058 (-0.145, 0.030) P=0.197		-0.055 (-0.163, 0.053) P=0.320	
Fast-food proximity				
Closer than 500m (ref)	1.00	0.89 (0.82, 0.97) P=0.006	1.00	0.92 (0.85, 0.99) P=0.029
500-999m	0.94 (0.88, 1.00) P=0.053	0.93 (0.86, 1.00) P=0.039	0.90 (0.83, 0.97) P=0.009	0.94 (0.88, 1.01) P=0.083
1000-1999m	0.98 (0.92, 1.04) P=0.472	0.90 (0.84, 0.97) P=0.008	0.99 (0.91, 1.08) P=0.843	0.92 (0.85, 0.98) P=0.017
At least 2000m	0.94 (0.88, 1.01) P=0.098	0.91 (0.84, 0.99) P=0.021	0.93 (0.84, 1.03) P=0.165	0.92 (0.85, 0.99) P=0.024
Stratum-specific HRs (≥2000m vs <500m)	0.97 (0.90, 1.05) P=0.451	0.97 (0.89, 1.07) P=0.591	0.93 (0.83, 1.04) P=0.187	1.00 (0.93, 1.07) P=0.966
Relative excess risk due to interaction (RERI)	0.082 (-0.012, 0.176) P=0.088		0.070 (-0.042, 0.182) P=0.218	
Greenspace				
Q1 (least greenspace) (ref)	1.00	0.86 (0.80, 0.93) P=0.000	1.00	0.93 (0.87, 1.01) P=0.070
Q2	0.98 (0.92, 1.05) P=0.595	0.91 (0.84, 0.98) P=0.012	0.99 (0.92, 1.07) P=0.764	0.95 (0.89, 1.01) P=0.129
Q3	0.96 (0.90, 1.03) P=0.249	0.94 (0.87, 1.01) P=0.112	1.00 (0.91, 1.10) P=0.976	0.95 (0.89, 1.01) P=0.094
Q4 (most greenspace)	0.95 (0.87, 1.02) P=0.166	0.96 (0.88, 1.04) P=0.287	0.86 (0.75, 0.98) P=0.024	0.96 (0.89, 1.04) P=0.325
Stratum-specific HRs (Q4 vs Q1)	0.97 (0.89, 1.06) P=0.490	1.05 (0.95, 1.17) P=0.336	0.84 (0.71, 0.98) P=0.027	1.04 (0.96, 1.13) P=0.332
Relative excess risk due to interaction (RERI)	0.149 (0.060, 0.238) P=0.001		0.170 (0.045, 0.296) P=0.008	

* Self-reported average total household income before tax

** 'More deprived' refers to people living in areas in the top two most deprived quintiles of the UK, based on the Townsend index.

Q = quartile

Combining area deprivation and household income, a beneficial association of having greater exposure to greenspace within 300 m of home was observed only among low-income households in deprived areas, where the hazard of cancer-related hospital admission was 24% lower among people living in the greenest quartile than among people living in the least green quartile (HR=0.76, 95% CI: 0.63-0.92, Table 8.6). Intermediate quartiles showed no significant difference from the least green quartile, but all HRs were less than one.

People from low-income households in deprived areas were also the group where living at least 2 km from a fast-food store seemed to have the greatest impact on cancer-related admissions (HR=0.88, 95% CI: 0.77-1.01), but there was no clear trend of decreasing hazard with decreasing proximity (Table 8.6).

No income/deprivation combined subgroup appeared to experience a substantial cancer-related benefit of having more PA facilities near home, although there was some evidence that those in low income households in more affluent areas had a somewhat reduced hazard (6%) if they had at least four PA facilities with a kilometre of home, compared with no facilities (HR=0.94, 95% CI: 0.86-1.02, Table 8.6).

c) Sex differences

For the relationships between all three neighbourhood exposures and cancer-related admission, the primary findings were generally consistent for women and men (Supplementary Tables 11-13, Appendix Five).

Table 8.6 Association between neighbourhood characteristics and cancer-related hospital admissions, stratified by household income and area deprivation in combination

Cancer-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
One	1.05 (0.93, 1.18) P=0.452	1.14 (0.93, 1.40) P=0.212	0.95 (0.88, 1.03) P=0.203	1.02 (0.94, 1.10) P=0.694
2-3	1.07 (0.95, 1.19) P=0.256	1.05 (0.87, 1.28) P=0.598	0.96 (0.89, 1.03) P=0.258	0.98 (0.90, 1.07) P=0.656
4 or more	1.05 (0.93, 1.18) P=0.434	0.98 (0.81, 1.19) P=0.850	0.94 (0.86, 1.02) P=0.152	0.96 (0.87, 1.05) P=0.324
Fast-food proximity				
Closer than 500m	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
500-999m	0.89 (0.81, 0.98) P=0.015	0.93 (0.80, 1.08) P=0.347	1.00 (0.92, 1.10) P=0.945	1.06 (0.96, 1.18) P=0.254
1000-1999m	0.97 (0.87, 1.08) P=0.569	1.01 (0.86, 1.20) P=0.899	1.03 (0.94, 1.12) P=0.572	0.98 (0.88, 1.09) P=0.733
At least 2000m	0.88 (0.77, 1.01) P=0.063	1.06 (0.85, 1.32) P=0.607	1.02 (0.93, 1.13) P=0.655	0.98 (0.88, 1.10) P=0.743
Greenspace				
Q1 (least greenspace)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Q2	0.98 (0.90, 1.08) P=0.720	0.97 (0.83, 1.13) P=0.654	1.02 (0.93, 1.13) P=0.612	1.03 (0.93, 1.15) P=0.566
Q3	0.96 (0.85, 1.08) P=0.456	1.03 (0.84, 1.26) P=0.768	1.01 (0.92, 1.12) P=0.803	1.05 (0.94, 1.18) P=0.357
Q4 (most greenspace)	0.76 (0.63, 0.92) P=0.005	1.05 (0.78, 1.40) P=0.752	1.04 (0.93, 1.16) P=0.543	1.08 (0.95, 1.22) P=0.233

* Self-reported average total household income before tax

** 'More deprived' refers to people living in areas in the top two most deprived quintiles of the UK, based on the Townsend index.

Q = quartile

d) Secondary outcomes: Breast and colorectal cancer

When we explored whether the results for cancer hospitalisations were driven by either of the two cancers most strongly linked to some of the plausible pathways by which neighbourhood characteristics might influence cancer risk (i.e. breast and colorectal cancer) we found that the evidence of effect modification by area deprivation of the association between greenspace and cancer was magnified for breast cancer (RERI=0.316, Supplementary Table 14) and the same was true for effect modification by household income (RERI=0.307). In deprived areas, the hazard of being admitted to hospital with a primary diagnosis of breast cancer was reduced by 31% among women with the greatest exposure to neighbourhood greenspace, compared with women who had the least greenspace near home (HR=0.69, 95% CI: 0.47 - 0.99, Supplementary Table 14). No such association was observed for women living in less deprived areas, and no association was observed between greenspace and breast cancer for the sample as a whole. For women from lower-income households who also lived in deprived areas, risk of a breast cancer-related admission was reduced by 39% among women with the greatest exposure to greenspace, compared with women who had the least exposure (HR=0.61, 95% CI: 0.38 - 0.97, Supplementary Table 15).

For formal PA facilities, no overall association was observed with either cancer subtype, just as was the case for all cancers combined. There was some indication of effect modification by income for breast cancer, such that women from higher income households may benefit from some protection against breast cancer if they have at least four PA facilities near home. Conversely, and consistent with the results for all cancer subtypes collectively, reduced colorectal cancer risk appeared to associated with greater availability of PA facilities among people living in more affluent areas, and in particular among people from lower-income households within more affluent areas (Supplementary Tables 16 & 17).

For fast-food proximity, neither cancer type showed an association with this neighbourhood exposure and there was limited evidence of any effect modification by either income or deprivation.

Sensitivity analyses

In general and for both outcomes, restricting follow-up to the period from 2012 onwards for all participants, rather than from the baseline assessment date, reduced precision around point estimates, but made minimal difference to the overall direction and magnitude of most coefficients and RERI estimates (Supplementary Tables 19–22). The main exception to this was that the departure from additivity due to area deprivation for the association between fast-food proximity and CVD-related admissions was far less when follow-up was restricted (RERI=-0.058 cf. RERI=-0.104). The RERI for effect modification by area deprivation of the association between PA facilities and CVD was also somewhat attenuated (RERI=0.070, cf. RERI=0.088), but in contrast effect modification of the same relationship by household income was amplified (RERI=-0.113 cf. RERI=-0.077), and the overall finding for the combination of income and deprivation was, if anything, stronger when follow-up was restricted to 2012 onwards. Sensitivity analyses adjusting for baseline BMI, hypertension and medications for cholesterol and hypertension, yielded very similar point estimates to the primary results (Supplementary Tables 23-26).

DISCUSSION

Across this very large sample of mid-aged adults in the UK, we examined the relationship between three characteristics of the neighbourhood built environment and hospital admissions due to CVD or cancer, over almost 10 years of follow up. We then examined whether these associations were modified by area deprivation and household income, with the aim of identifying which neighbourhood characteristics might best be intervened on to improve health without widening existing health inequalities.

For the sample as a whole, we observed a weak trend of reducing hazard of hospital admission due to CVD with increasing distance to the nearest takeaway/fast-food store, and some protection for people with the greatest availability of PA facilities within one kilometre of home; however, the 95% confidence intervals did not exclude the null value of no hazard reduction. No such association was apparent between neighbourhood greenspace coverage and CVD, and we observed very little evidence that any of the three neighbourhood exposures were associated with hospitalisations due to cancer in the sample as a whole.

More noteworthy, however, are the findings for effect modification by household income and area deprivation, where we observed some interesting patterns that may help to illuminate important elements of the links between the neighbourhood built environment

and health. The largely null associations in the sample as a whole appeared to be masking potentially important variation in the strength and magnitude of some of those associations according to socioeconomic conditions.

CVD-related hospital admissions

For the availability of formal PA facilities and CVD outcomes, we observed evidence of effect modification on the additive scale by area deprivation, suggesting that intervening to improve access to formal PA facilities in deprived areas is likely to have a greater public health impact than it would in less deprived areas. At the same time, we also observed evidence of effect modification on the additive scale by household income, but in this case the association was stronger among higher-income households (after controlling for area deprivation). Although the magnitude of the departure from additivity was slightly smaller than that observed in the other direction for area deprivation, and the 95% CIs for both RERIs include zero, the direction of the RERI for income is consistent with the fact that most formal PA facilities impose some financial cost on users. We would therefore be surprised if we did not observe the hypothesised health benefits of greater neighbourhood availability of these facilities accruing disproportionally to higher-income households. The contrasting directions of the additive measures of effect modification by (higher) income and (less) deprivation imply that greater availability of formal PA facilities is particularly beneficial in deprived areas, but only for the those who can afford to access those facilities. Consistent with this we found that when we considered household income and area deprivation together, the estimated benefits of greater availability of neighbourhood PA facilities was indeed largely restricted to high-income households in deprived areas, among whom we observed a one-fifth reduction in the hazard of being admitted to hospital with CVD for people living near at least four PA facilities, compared with those people who had no local PA facilities. If the CVD-related benefits of greater availability of PA facilities accrue primarily to high-income households in deprived areas, the policy implications are obvious: locating more PA facilities in deprived areas may reduce CVD risk, but the greatest gains stand to be made if facilities in those areas are accessible to all, regardless of income. Otherwise, health benefits in deprived areas may be concentrated among the well-off living there, thus widening health inequalities. A recent quasi-experimental study in a deprived local authority in the north west of England[32] showed improved outcomes following the introduction of universal free access to council leisure facilities, indicating that this is indeed a potentially effective approach to adopt.

With respect to fast-food proximity and risk of CVD-related hospital admission, again the largely null main associations appear to conceal stronger associations in some subgroups. The findings for effect modification by income and area deprivation on the additive scale suggest that reducing access to fast-food stores may have the greatest impact for low-income households, especially men in those households, but mostly in relatively affluent areas. The RERI for area deprivation as a modifier of the association with CVD is less than zero, and stratified analysis shows that greater distances from home to the nearest fast-food store are more strongly associated with reduced risk of CVD-related hospital admissions among people living in less deprived areas. This may imply a greater range of alternatives to fast food available to residents of more affluent areas. If so, this supports other evidence for the importance of considering the whole retail food environment – healthy and unhealthy stores – when making planning decisions. And as was seen for PA facilities, there was evidence of a contrasting role of income: but in this case the impact of fast-food proximity seems to be stronger among people in low-income households. The latter finding is consistent with as similar recent study using UK Biobank but concentrating only on participants in London[15] and makes sense because lower income households are likely to be more sensitive to price and perceived value for money offered by fast-food stores, and may also rely more on small, frequent shopping trips close to home rather than large weekly shopping trips to a large supermarket further away, and on the convenience of food outlets near home[33].

Unlike for fast-food proximity and availability of formal PA facilities, there was little evidence of an association between neighbourhood greenspace – operationalised as domestic gardens and public greenspace with 300 m of home – and CVD-related hospital admissions for any income or deprivation subgroup. Instead, the null association observed for the sample as a whole was largely preserved across household income and area deprivation levels, with no evidence of effect modification. The observation of area deprivation driving a positive departure from additivity for formal PA facilities but not for greenspace may reflect poorer quality greenspace[34] or perceived safety concerns in deprived areas[17], or other factors making greenspaces less suitable for PA. It is worth noting that the number of observations in the most deprived, most green combination was small, leading to a lack of precision around that hazard ratio, and correspondingly around the RERI. Thus, we should be cautious about drawing conclusions about whether or not area deprivation modifies the estimated effect of greenspace on CVD risk.

Cancer-related hospital admissions

With respect to all cancer-related hospital admissions, we found no evidence of any association between formal PA facilities and cancer, overall or within any income or area deprivation subgroup (and conflicting patterns of effect modification for breast and colorectal cancer). In contrast, there did appear to be an association between neighbourhood greenspace and cancer for people living in more deprived areas, which was masked when looking at the sample as a whole. This was especially the case for breast cancer-related admissions. The strong evidence of effect modification by area deprivation was the largest we observed and indicates a greater protective influence of more greenspace against cancer in more deprived areas than in less deprived areas. This finding is consistent with other studies that have previously found that relationships between greenspace and health appear to be stronger in more deprived communities in the UK[35,36].

One pathway through which greenspace is hypothesised to influence health is via physical activity. The fact that in deprived areas we observed a relationship between greenspace and cancer but no relationship between greenspace and CVD is therefore interesting, and suggests greenspace might influence cancer risk through pathways unrelated to PA. This is backed up by the lack of evidence in this study of an association between PA facilities and cancer, including in deprived areas. That we see more evidence of an association between PA facilities and CVD in deprived areas but more evidence of an association between greenspace and cancer in deprived areas, may indicate that these two neighbourhood resources influence health via different pathways. While formal PA facilities are unlikely to influence health via pathways other than through physical activity itself, there is emerging evidence that greenspace may influence health via multiple pathways, including mental wellbeing, immune function, and respiratory health[13], as well as PA. Several studies have concluded that greenspace-health relationships, if causal, are mediated by pathways other than PA, most notably psychosocial ones[37–39]. One of the principal mechanisms by which greenspace is thought to influence health is the regulation of cortisol secretion[40]. Cortisol secretion is an indicator of stress and its dysregulation is associated with various health outcomes including both cancer and CVD risk[41]. A recent study in a deprived setting in Scotland found that the presence of more greenspace near the home was associated with lower levels of stress across objective cortisol secretion measures and subjective measures of stress, but this relationship did not appear to be mediated by physical activity[40]. Access to greenspace near home may also plausibly mitigate other biological pathways through which chronic psychological stress (more prevalent in deprived populations) influences cancer risk, such as oxidative stress-induced

DNA damage and telomere shortening[42,43]. Similarly, greenspace may mitigate some of the effects on cancer risk of air and noise pollution (also often higher in deprived areas), operating through these and related inflammatory and oxidative stress pathways[44,45].

For fast-food proximity and cancer, there was no evidence of an interaction with income, but some evidence that area deprivation modifies the effect of fast-food proximity, although in the opposite direction to what we observed for CVD. The measure of fast-food proximity we have used is somewhat problematic, however, and these results may not be reliable for either outcome. An ideal measure would capture both proximity and density of facilities in an area, and take into account the wider neighbourhood food environment, in terms of access to both healthy and unhealthy food, which are often highly correlated[46]. There is likely to be some systematic misclassification, random error, and geographical inconsistency in quality in the proximity measure we have used, due to our reliance on an off-the shelf measure based on local authority data sources collected for non-research purposes. This highlights some of the trade-offs made in the use of big data and administrative data for the purposes of epidemiological research. Further research repeating this UK-wide analysis using improved measures of the fast-food environment may clarify this relationship.

Our findings for the overall associations between these neighbourhood exposures and CVD and cancer are generally small in magnitude and in most cases null. An important *a priori* rationale for examining effect modification by factors such as income and area deprivation, when a study is sufficiently powered to do so, is that it is plausible that some groups of people will be more sensitive to their neighbourhood environment than others, and that some may be almost completely insensitive for various reasons. Population-wide, average effect estimates smooth out these differences and potentially lead to erroneous conclusions about the importance of neighbourhood environments for some people in some places. We would only expect small effect sizes, given the complexity and multitude of causes of these health outcomes, and how distal they are from the exposures, but in some cases the null or very weak findings contradict what we might expect based on previous research. In particular, evidence from food environment research in the UK has been mounting of a detrimental effect of excessive exposure to unhealthy food outlets[15,47–49]. Limitations of the fast-food proximity measure are described above, and are also likely to have led to conservative estimates. Similarly, the greenspace measure may also not adequately capture the full extent of relevant greenness of one's neighbourhood, as it does not include smaller parcels of greenspace such a street trees, or reflect 'quality' of greenspace. Related to this, because the greenspace measure is based on 2005 data, it may

better reflect the true exposure to neighbourhood greenspace for those participants recruited earlier in the baseline phase than those recruited later.

There are several other limitations of the current study. First, the hospital admissions data only captures inpatient care, so any early detection of CVD and cancer that occurs in primary care settings after baseline and is then effectively treated without admission to hospital will not be counted. Such cases are probably more likely to occur in higher income or less deprived subgroups[50], and this may have contributed to lower risk of the outcome in those groups, potentially distorting the magnitude of effect modification on the additive scale. In the future, when GP records are fully linked to the UK Biobank cohort, it will be possible to examine this potential source of bias. Related to this, if some types of health care have shifted to outpatient settings over the course of the follow-up period, it may result in some dilution of the true association overall and between subgroups.

Second, it is unclear what period of follow-up is likely to be necessary to capture the effect of interest, given that people will have been exposed to their baseline neighbourhood conditions for varying lengths of time depending on how long they have lived at that address, and whether relevant changes had occurred in their neighbourhood during that time, and the nature of previous neighbourhood exposures. We adjusted our analyses for years living at baseline address to attempt to deal with this, and are reassured by the long average time people have lived at the address we are using (median=15 years), but there may be remaining imprecision, and potential bias of estimates in either direction, that we cannot overcome using observational data of this kind. Longer follow up may prove to be more revealing, and that will become possible in future years, but ideally future work would also account for changes in the built environment over that period. UK Biobank would be made richer by the addition of measurement of neighbourhood exposures at one or more post-baseline time points. Our sensitivity analyses using a shorter follow-up period to account for the timing of the exposure ascertainment showed that most point estimates were robust to this specification, but there was a loss of precision presumably driven by the substantial reduction in the number of hospital admissions occurring during the shortened follow-up period (Supplementary Table 18).

Finally, we cannot rule out self-selection into more health promoting neighbourhoods by people more disposed to healthy behaviours. We can, however, by the longitudinal nature of the study and exclusion of people with prevalent disease at baseline, rule out active self-selection prior to baseline into neighbourhoods on the basis of prevalent disease (e.g. following a cardiac event earlier in life, deciding to relocate to a neighbourhood more

supportive of a healthy lifestyle). This means that we likely minimise masking of the true effect via this avenue, but may still have some residual positive confounding that could bias the association away from the null, despite our comprehensive adjustment for observed potential confounders. However, UK Biobank is a residentially very stable sample, and most of our strongest findings were within more deprived subgroups, where financial resources enabling relocation for health purposes are presumably the least.

Overall, despite no estimated protective effect of greenspace on cancer and CVD across the mid-aged English population taken as a whole, subgroup effects were observed. Living in a neighbourhood with a greater percentage of greenspace is associated with lower risk of cancer-related hospitalisation among people living in more deprived areas. There is some evidence of the same being true for reduced fast-food proximity and cancer, though the opposite was observed for fast-food proximity and CVD. Greater availability of PA facilities close to home is associated with lower risk of CVD admissions in more deprived areas, but also only among those with higher household incomes. Improving deprived neighbourhoods by increasing the number of formal PA facilities, while also ensuring access to these is free or affordable, and by increasing the amount of public and private greenspace and limiting the proximity of fast-food outlets to residential areas, may improve health outcomes in the population.

Taken together, these results suggest that improving access to both greenspace and PA facilities may have a greater public health impact in more deprived areas, but the pathways by which these benefits might arise require further elucidation and should not be assumed to be restricted to the promotion and facilitation of physical activity. In other words, increasing access to both is likely to be more beneficial to health overall than focusing on one or the other. We also show that by examining effect modification by multiple socioeconomic indicators in parallel, potentially important insights can be gained that may be missed when we focus only on a single measure of either household or area-level socioeconomic conditions. Understanding the potentially different ways in which different aspects of the socioeconomic conditions of people's lives influence their relationship with the built environment and its effects on their health may help to avoid intervention-generated inequalities when neighbourhood-based built environment interventions are designed.

REFERENCES

1 Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *J Public Health* 2011; **33**: 527–35.

2 Steptoe A, Kivimäki M. Stress and cardiovascular disease. *Nat Rev Cardiol* 2012; **9**: 360–70.

3 Allender S, Rayner M. The burden of overweight and obesity-related ill health in the UK. *Obes Rev* 2007; **8**: 467–73.

4 Secretan BL, Ph D, Scoccianti C, *et al.* Body fatness and cancer: viewpoint of the IARC working group. *N Engl J Med* 2016; **375**: 794–8.

5 Malambo P, Kengne AP, De Villiers A, Lambert E V., Puoane T. Built environment, selected risk factors and major cardiovascular disease outcomes: a systematic review. *PLoS One* 2016; **11**: 1–13.

6 Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev* 2011; **12**: 217–30.

7 Mackenbach JD, Rutter H, Compernolle S, et al. Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. *BMC Public Health* 2014; **14**: 233.

8 Gascon M, Mas MT, Martínez D, et al. Mental health benefits of long-term exposure to residential green and blue spaces: a systematic review. Int J Environ Res Public Health 2015; 12: 4354–79.

9 McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int J Behav Nutr Phys Act* 2011; **8**: 125.

10 Giskes K, van Lenthe F, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev* 2011; **12**: e95–106.

11 Alcock I, White MP, Wheeler BW, Fleming LE, Depledge MH. Longitudinal effects on mental health of moving to greener and less green urban areas. *Environ Sci Technol* 2014; **48**: 1247–55.

12 Feng X, Astell-Burt T, Badland H, Mavoa S, Giles-Corti B. Modest ratios of fast food outlets to supermarkets and green grocers are associated with higher body mass index: longitudinal analysis of a sample of 15,229 Australians aged 45 years and older in the Australian National Liveability Study. *Heal Place* 2018; **49**: 101–10.

13 World Health Organization. Urban green spaces and health: a review of the evidence. 2016.

14 Jones-Smith JC, Karter AJ, Warton EM, *et al.* Obesity and the food environment: income and ethnicity differences among people with diabetes: the Diabetes Study of Northern California (DISTANCE). *Diabetes Care* 2013; **36**: 2697–705.

15 Burgoine T, Sarkar C, Webster CJ, Monsivais P. Examining the interaction of fast-food outlet exposure and income on diet and obesity: evidence from 51,361 UK Biobank participants. *Int J Behav Nutr Phys Act* 2018; **15**: 1–12.

16 Lovasi GS, Neckerman KM, Quinn JW, Weiss CC, Rundle A. Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index. *Am J Public Health* 2009; **99**: 279–84.

17 Hobbs M, Green MA, Griffiths C, et al. Access and quality of parks and associations with obesity: a cross-sectional study. *SSM - Popul Heal* 2017; **3**: 722–9.

18 Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *Int J Behav Nutr Phys Act* 2007; **4:** 32.

19 Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013; **67**: 190–3.

20 UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007.

21 Sarkar C, Webster C, Gallacher J. UK Biobank Urban Morphometric Platform (UKBUMP): a nationwide resource for evidence-based healthy city planning and public health interventions. *Ann GIS* 2015; **21**: 135–48.

22 Wheeler B. Environmental indicators attributed to participant's home location (UK Biobank Project 15374). 2016.

23 Leitzmann M, Powers H, Anderson AS, *et al.* European code against cancer 4th edition: physical activity and cancer. *Cancer Epidemiol* 2015; **39S**: S46-55.

24 Norat T, Scoccianti C, Boutron-Ruault MC, *et al.* European code against cancer 4th edition: diet and cancer. *Cancer Epidemiol* 2015; **39S**: S56–66.

25 Vanderweele T. Explanation in Causal Inference: Methods for Mediation and Interaction. New York : Oxford University Press, 2015.

26 Grahn P, Stigsdotter UA. Landscape planning and stress. *Urban For Urban Green* 2003; **2**: 1–18.

27 Coles RW, Bussey SC. Urban forest landscapes in the UK: progressing the social agenda. *Landsc Urban Plan* 2000; **52**: 181–8.

28 Natural England. Nature Nearby - Accessible Natural Greenspace Guidance. 2010.

29 Davey Smith G, Hart C, Watt G, Hole D, Hawthorne V. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study. *J Epidemiol Community Heal* 1998; **52**: 399–405.

30 Vandenbroucke JP, von Elm E, Altman DG, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; **4**: e297.

31 Li R, Chambless L. Test for additive interaction in proportional hazards models. *Ann Epidemiol* 2007; **17**: 227–36.

32 Higgerson J, Halliday E, Ortiz-Nunez A, Brown R, Barr B. Impact of free access to leisure facilities and community outreach on inequalities in physical activity: a quasi-experimental study. *J Epidemiol Community Heal* 2018; **0**: 1–7.

33 Harrington J, Friel S, Thunhurst C, Kirby A, McElroy B. Obesogenic island: the financial burden of private transport on low-income households. *J Public Health* 2008; **30**: 38–44.

34 CABE. Urban Green Nation: Building the Evidence Base. London, UK, 2010.

35 Mitchell R, Popham F. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet* 2008; **372**: 1655–60.

36 Roe J, Aspinall AP, Ward Thompson C. Understanding relationships between health, ethnicity, place and the role of urban green space in deprived urban communities. *Int J Environ Res Public Health* 2016; **13**: 681.

37 James P, Hart JE, Banay RF, Laden F. Exposure to greenness and mortality in a nationwide prospective cohort study of women. *Environ Health Perspect* 2016; **124**: 1344–52.

38 Lachowycz K, Jones AP. Does walking explain associations between access to greenspace andlower mortality? *Soc Sci Med* 2014; **107**: 9–17.

39 De Vries S, van Dillen SME, Groenewegen PP, Spreeuwenberg P. Streetscape greenery and health: stress, social cohesion and physical activity as mediators. *Soc Sci Med* 2013; **94**: 26–33.

40 Ward Thompson C, Roe J, Aspinall P, Mitchell R, Clow A, Miller D. More green space is linked to less stress in deprived communities: evidence from salivary cortisol patterns. *Landsc Urban Plan* 2012; **105**: 221–9.

41 Nader N, Chrousos GP, Kino T. Interactions of the circadian CLOCK system and the HPA axis. *Trends Endocrinol Metab* 2010; **21**: 277–86.

42 Willeit P, Willeit J, Mayr A, *et al.* Telomere length and risk of incident cancer and cancer mortality. *JAMA* 2010; **304**: 69–75.

43 Martens DS, Nawrot TS. Ageing at the level of telomeres in association to residential landscape and air pollution at home and work: a review of the current evidence. *Toxicol Lett* 2018; **298**: 42–52.

44 Roswall N, Raaschou-Nielsen O, Ketzel M, Overvad K, Halkjær J, Sørensen M. Modeled traffic noise at the residence and colorectal cancer incidence: a cohort study. *Cancer Causes Control* 2017; **28**: 745–53.

45 Andersen ZJ, Stafoggia M, Weinmayr G, *et al.* Long-term exposure to ambient air pollution and incidence of postmenopausal breast cancer in 15 European cohorts within the ESCAPE project. *Environ Health Perspect* 2017; **125**: 107005.

46 Clary CM, Ramos Y, Shareck M, Kestens Y. Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. *Prev Med (Baltim)* 2015; **71**: 83–7.

47 Green MA, Radley D, Lomax N, Morris MA, Griffiths C. Is adolescent body mass index and waist circumference associated with the food environments surrounding schools and homes? A longitudinal analysis. *BMC Public Health* 2018; **18**: 482.

48 Burgoine T, Forouhi NG, Griffin SJ, Brage S, Wareham NJ, Monsivais P. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study. *Am J Clin Nutr* 2016; **103**: 1540–7.

49 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 2014; **348**: g1464.

50 Richard C, Carol P, Miqdad A, Rosalind R. Socio-Economic inequalities in health care in England. *Fisc Stud* 2016; **37**: 371–403.

Chapter 9. DISCUSSION & CONCLUSIONS

In this concluding chapter I summarise and synthesise the key findings, and outline the strengths and limitations of the project as a whole. I conclude with a discussion of the overall contribution of the work to the field and discuss policy implications and directions for future research that emerge from this work.

9.1. Summary of study rationale, aims and objectives

9.1.1. Rationale and aims

Socio-ecological models of health posit that determinants of health operate at various scales, from the individual through to the macro-environmental[1,2]. Local residential areas, or neighbourhoods, are one such scale[3]. Various built environment resources in neighbourhoods are likely to influence health, and neighbourhoods are a potentially valuable site for public health interventions. Yet the evidence base regarding neighbourhood effects on health remains equivocal. A possible explanation for this is that some people may be more sensitive than others to their neighbourhood built environment, and neighbourhoods may matter more for health in some places than in others – in other words, neighbourhood effects may be heterogeneous rather than universal[3]. Furthermore, although socio-ecological models of health conceptualise determinants of health operating at multiple levels, interactions between neighbourhood characteristics and factors operating at other scales remain underexplored[4].

In this thesis I have therefore undertaken a multi-scalar examination of sources of potential effect heterogeneity in neighbourhood-health relationships. In doing so I have endeavoured to contribute to our understanding of how, where and for whom the neighbourhood built environment matters for health. Improving this understanding may serve two purposes: (1) yielding new insights into the ways neighbourhood environments contribute to a complex, multi-level system of influences, operating together to produce complex health outcomes; and (2) helping to guide and target future interventions to maximise public health impact.

Specifically, in a very large sample of UK adults at a critical stage of the life course, I investigated associations between multiple characteristics of the neighbourhood built environment and outcomes relating to obesity, cardiovascular disease and cancer, with a focus on characterising potential sources of effect heterogeneity operating at the individual, neighbourhood and macro-environmental level. The breadth and depth of the UK Biobank resource, including the associated UKBUMP database and other

environmental data, linked health records, and comprehensive genotyping of individuals, enabled robust analysis of numerous hypothesised effect modification relationships, in addition to an examination of the primary associations of interest.

Good epidemiological practice avoids *post hoc* subgroup analysis, so it is important that research focussing on effect heterogeneity is theoretically grounded and hypothesis driven. Therefore, after the initial paper in which I examined main associations averaged across the population, I designed the rest of this study around a series of specific effect modification hypotheses, some of which already had some support in the literature, and others that had not previously been tested, but which were biologically plausible and identified in conceptual models or theoretical discussions. In the introduction to this thesis I presented a conceptual model (Figure 1.1, p.19), adapted from similar ones by Kremers[5] and Schneider[6] and drawing on the ANGELO framework[7]. It shows the pathways from neighbourhood exposures to health as potentially being modified by other factors operating at various scales. This model provided a framework for the hypotheses I have tested in the five research papers in this thesis, addressing potential modification of neighbourhood effects on important health outcomes by individual, neighbourhood and macro-environment variables.

9.1.2. Study objectives

Using observational data from UK Biobank – a uniquely large sample of adults in a critical period of the life course for the development of chronic disease – I sought to address the following objectives:

1. To assess whether characteristics of the food and physical activity environments near an individual's place of residence are independently associated with objectively measured adiposity (BMI, waist circumference and % body fat), and whether this varied by sex and household income. (Chapter 4)
2. To assess whether genetic risk of obesity modifies associations between neighbourhood characteristics and body mass index. (Chapter 5)
3. To assess whether the association between availability of formal physical activity facilities and adiposity is modified by other features of the neighbourhood environment, namely proximity to fast-food outlets and availability of public green spaces. (Chapter 6)
4. To examine possible geographical heterogeneity in the associations between the neighbourhood physical activity and fast-food environments and BMI across England, and explore whether any such heterogeneity might be explained by locally varying factors. (Chapter 7)

5. To assess whether characteristics of neighbourhood environments are associated with being admitted to hospital with a primary diagnosis of CVD or cancer, and whether these associations are modified by household income and area deprivation. (Chapter 8)

Each objective was addressed in a separate empirical chapter, centred around a research paper (either published, under review or ready for submission to a journal). Across the study, I used a mix of regression modelling techniques, informed by careful consideration of potential sources of bias in the estimation of the associations of interest. Secondary objectives of the thesis were therefore to examine how robust the main findings were to some of the methodological choices I made, and whether the main results were likely to be affected by bias, especially residual confounding. In each chapter I presented various sensitivity analyses that I designed to meet these objectives.

In addressing objectives 1-4, I used cross-sectional data from the UK Biobank baseline study, and focussed on adiposity measures as important intermediate health outcomes for which there are clear plausible pathways to health from what are often referred to as 'obesogenic' neighbourhood characteristics. If food and physical activity environments influence dietary and physical activity behaviours, this should in turn be reflected in BMI and other measures of adiposity such as waist circumference and percent body fat. Then, because excess adiposity (i.e. overweight and obesity) is an important risk factor for many non-communicable diseases, objective 5 expands the earlier focus to include more distal health outcomes, making use of linked administrative health records to examine incident CVD and cancer over a ten-year follow up period. For these more distal outcomes I also considered that neighbourhood environments may exert an influence through pathways other than diet and physical activity, so I additionally examined neighbourhood greenspace as an exposure.

9.2. Summary of findings

Chapter 1 of the thesis provided a background to the project and outlined the aims and objectives. In Chapters 2 and 3 I introduced the data and methods I have used. In Chapter 4 I began the empirical work of the thesis by establishing the independent associations between adiposity and two residential environmental exposures likely to influence adiposity via distinct pathways: fast-food proximity, via energy intake; and availability of PA facilities, via energy expenditure. Following that, the remaining empirical work of the thesis presents a systematic, theoretically grounded exploration of potential heterogeneity of these and related relationships, considering modifiers at multiple scales, from the individual (genetic, socioeconomic) to the neighbourhood (physical and social

environments), and the wider contexts in which neighbourhoods are nested (physical and socio-cultural attributes of local authority districts).

In summary, the main findings of the five research papers that form the basis of Chapters 4-8 were as follows:

Characteristics of the neighbourhood food and physical activity environment are independently associated with objectively measured adiposity. In Chapter 4 I found that local access to formal PA facilities such as leisure centres, gyms and sports fields is independently associated with adiposity. As the density of formal PA facilities increased, BMI, waist circumference and percentage body fat all decreased. A similar but weaker association was observed between proximity to fast-food/takeaway outlets and the same adiposity measures. There was strong evidence of effect modification by sex and income, with stratified models showing modestly larger estimates of effects of both neighbourhood exposures on women's adiposity than on men's, and larger effects of formal PA facility availability among people from higher-income households. No association was observed between neighbourhood greenspace and adiposity.

Individuals at increased genetic risk of obesity may be more sensitive to their local food environment, but no gene-environment interaction was observed for the physical activity environment. In Chapter 5 I considered genetic risk of obesity as a potential individual-level modifier of associations between BMI and the two neighbourhood exposures examined in Chapter 4. This made use of the genotype data available for the UK Biobank cohort, to explore a novel research question. While I found very limited evidence that genetic risk of obesity modifies the association between availability of physical activity facilities and BMI, there was good evidence of a gene-environment interaction for proximity to a fast-food/takeaway outlet. The association between fast-food proximity and BMI was stronger among those at increased genetic risk of obesity, both when genetic risk was measured by polygenic risk scores, and in particular for a single genetic marker linked to *MC4R*, a gene known to regulate food intake. This suggests individuals at increased genetic risk of obesity may be more sensitive to their local food environment.

Other resources in the neighbourhood built environment modify the association between the availability of formal PA facilities and adiposity. In Chapter 6 I hypothesised that the negative association between the neighbourhood availability of formal PA facilities and BMI observed in Chapter 4 would vary according to the fast-food environment (which might negate any beneficial impacts of a healthy PA environment) and the informal physical activity environment (operationalised as availability of parks and other public spaces near home, which provide alternative sites for PA). Indeed, the association was

somewhat stronger among people with fewer urban parks and other open/green spaces in their neighbourhood than among those with more of these resources, and was noticeably attenuated among those who lived closer (<500 m) to a fast-food store, compared with people living further away. This suggests formal physical activity facilities may buffer against a lack of informal, green resources for physical activity in areas where the latter are scarce. However, the potential benefits of physical activity facilities in terms of obesity risk may be undermined by an unhealthy food environment close to home.

There is considerable geographical heterogeneity across England in the magnitude and direction of the relationship between neighbourhood PA and fast-food environments and BMI, and locally varying factors of the wider context may explain some of this heterogeneity. In Chapter 7 I widened the scale and considered effect modifiers operating at the level of the local authority, having first described geographical heterogeneity at that level. To facilitate investigation of these cross-level interactions I geolocated the UK Biobank participants in England in their respective Local Authority Districts and then linked multiple externally sourced datasets that described physical and socio-cultural attributes of the local authority districts. The association between availability of neighbourhood PA facilities and BMI showed considerable heterogeneity between local authorities across urban England, from strong associations in some districts through to no association in others. The same was true for the association between fast-food proximity and BMI. This may have implications for the generalisability of studies with a narrow geographical focus. Although differences weren't large, the association between the formal PA environment and BMI was weak among people in local authorities with the most natural land cover, but stronger among people living in local authorities with less natural land cover, possibly indicating a greater reliance on or normalisation of the use of formal physical activity facilities there. Evidence of this effect modification relationship was more apparent outside London.

Potentially important differences between socioeconomic subgroups are obscured when examining only population-wide estimates of the effects of neighbourhood characteristics of CVD and cancer. In Chapter 8 I extended my focus from adiposity outcomes to hospital admissions caused by two of the NCDs for which adiposity is a major risk factor: CVD and cancer. I assessed effect modification of these relationships by household income and area deprivation. I also examined neighbourhood greenspace as a third potentially important exposure. I found that the association between availability of PA facilities and CVD (but not cancer) outcomes appears to be modified on the additive scale by both area deprivation and household income, suggesting that intervening to improve access to PA

facilities is likely to have a greater public health impact in more deprived areas, but also that higher-income households are more likely to benefit than low-income households. I also observed evidence of an additive interaction between area deprivation and neighbourhood greenspace in relation to cancer (but not CVD) such that greenspace appears to be more protective against cancer in more deprived areas than in less deprived areas. These findings suggest that in populations where the formal PA environment and greenspace influence health, they may do so via different pathways. These subgroup-specific relationships were obscured in analyses that ignored possible interactions with socioeconomic circumstances. In those models without interactions, I observed only very weak associations between fast-food proximity and both CVD and cancer hospital admissions and between availability of PA facilities and cancer, and no evidence of any associations between neighbourhood greenspace and either outcome.

9.3. Synthesis of findings

Overall, the findings of this thesis highlight the importance of being attentive to possibly heterogeneous effects, and of considering any given neighbourhood exposure or risk factor within its broader context. Where evidence for beneficial effects of a neighbourhood characteristic exists, we should infer that those benefits may not be distributed equally, and recognise the implications of that for intervention. Where effect estimates are small or consistent with no effect, global averages may be masking important subgroup effects, and thereby concealing potentially informative evidence about the underlying mechanisms by which neighbourhoods influence health.

As mentioned earlier in this chapter and discussed in Chapter 1, the rationale for studying effect heterogeneity and its sources in this thesis (and in general) is twofold: (1) to gain a deeper aetiological understanding of how factors at multiple levels interact to shape health, and (2) to help determine which subpopulations would benefit most from intervention – an issue of considerable public health importance. Subject to confirmatory studies, the work of this thesis contributes to both these ends. Here, I synthesise the key findings of the project as they relate to these two purposes, drawing out several cross-cutting themes that emerged in relation to each purpose.

9.3.1. Exploring effect heterogeneity to improve understanding of health and disease

9.3.1.1. Effect heterogeneity underlines the complexity of relationships between neighbourhood environments and health

It is increasingly being acknowledged that upstream factors are key to today's high prevalence of obesity and non-communicable diseases such as CVD and cancer, and that the established narrative of individual responsibility ignores the reality that individual behavioural choices are made in a context of ready availability of unhealthy food options, physical barriers and structural disincentives to being physically active, pervasive industry marketing, financial pressures and social norms[8]. In addition, these behavioural choices are made against a backdrop of individual risk factors (e.g. genetic risk) and personal circumstances (e.g. socioeconomic position) over which individuals have little or no control.

The combination of multi-layered environmental and social context, and individual risk, creates complex systems of the kind comprehensively illustrated in the UK's Foresight report in 2007[9]. While much research has been and continues to be done to understand the role of separate components of these systems, we are a long way from understanding the dynamics of these systems, the synergisms and antagonisms within them, their feedback loops and their emergent properties[10]. I have not gone so far as to adopt a complex systems modelling approach in this thesis; rather I have sought to use traditional social epidemiological methods to understand some of the interacting relationships within those systems. This approach has been called for in the social epidemiology and neighbourhood effects literatures in recent years[4,11–13], in the hope that it may help to make sense of the many inconsistent findings in the literature.

The fact that in research to date we do not consistently observe strong associations in the expected directions for any characteristics of the neighbourhood built environment, coupled with an abundance of null findings, might lead us to conclude neighbourhood environments don't matter after all. A more compelling conclusion, however, is that because these environments are part of a complex, multi-level system of influences operating in different settings and over the whole of the lifecourse, they are unlikely to have a uniform effect across the population. Each empirical chapter within this thesis provides evidence to support this conclusion. Building on existing literature, I report evidence of relationships between neighbourhood characteristics and either adiposity or NCDs, that vary across population subgroups or geographical space, with effect modifiers operating at scales from the genetic right through to the macro-environment. Further

discerning *for whom* and *where* specific neighbourhood characteristics matter remains of critical importance for the understanding of *how* neighbourhoods influence health.

9.3.1.2. Individual-level factors moderate sensitivity to the health-promoting and health-damaging influences of neighbourhood environments

Various characteristics of individuals might plausibly interact with neighbourhood factors to influence obesity and NCD risk, and understanding these interactions can shed new light on how the built environment becomes, as Krieger puts it, embodied[14]. Yet research to date has concentrated on only a small subset of these. One important but underexamined variable is genetic risk. Contributing to an improved understanding of *how* neighbourhoods influence health, in Chapter 5 I found that individuals at increased genetic risk of obesity may be more sensitive to exposure to fast-food outlets. In particular, the strong evidence of an interaction between fast-food proximity and a specific SNP near *MC4R*, a gene known to be involved in regulation of food intake – but no such interaction between fast-food proximity and markers associated with physical activity, or between the *MC4R* marker and the PA environment – lends support to the biological plausibility of this potentially important gene-environment interaction. If unhealthy features of the neighbourhood food environment are confirmed to pose a greater obesity risk to those already genetically predisposed to higher BMI, this will shed light on the ways that genetic risk and neighbourhood environments individually and in concert contribute to obesity in the population. The novel evidence provided here on this question will hopefully prompt confirmatory studies and further exploration and elucidation of this relationship. The findings are consistent with other emerging evidence that the contexts in which we develop and live shape the way genetic factors influence our health. For example, one recent study reported that cohort of birth modified the effect of the obesity-associated *FTO* gene[15], and another reported that perceptions of the neighbourhood environment may exacerbate genetic risk of diabetes[16]. Taken together, these findings suggest that GxE interactions of the kind I have explored here may exist across a range of contextual exposures and a range of health outcomes.

Already relatively widely studied, gender differences in neighbourhood effects are another source of effect heterogeneity that offers insight into the mechanisms by which specific neighbourhood characteristics influence particular health outcomes. Confirming what some researchers have reported previously in the UK and elsewhere[17,18], but contrasting with some other UK studies[19,20], in both Chapter 4 and Chapter 8 I observed differences between women and men in the magnitude of associations. In Chapter 4, neighbourhood-adiposity associations were of greater magnitude among women, especially with respect to PA facilities. In Chapter 8, associations with CVD and cancer also showed some

variation by sex, as did whether or not these associations were modified by socioeconomic factors. Whether men or women were most affected depended on which exposure was being considered. This makes sense, as for each exposure there will be distinct processes at work in the production of differential effects in women and men. CVD outcomes among low-income men appeared most sensitive to fast-food proximity, while cancer outcomes among women living in deprived areas were most sensitive to neighbourhood greenspace. If causal, these hint at gendered relationships with place. Qualitative research to unpack the different ways (some) men and (some) women engage with, perceive, act within, and are constrained or enabled by, neighbourhood environments, is likely to be informative. For example, Coen et al[21] recently used qualitative methods to examine the gendering of physical activity in Canadian gyms, highlighting numerous ways in which gender influences physical activity .While evidence of sex and gender differences within this study and across other studies is mixed, it does suggest a need to take sex and gender into account in research on environmental determinants of obesity and related outcomes, and highlights the likely importance of seeking and applying a more nuanced understanding of how men and women interact with and experience their neighbourhoods differently.

9.3.1.3. Both individual- and area-level socioeconomic position can interact with the neighbourhood built environment to influence health, though not always in the same ways

It is already well established that socioeconomic circumstances play an important role in determining health outcomes[22]. Indeed, this is central to the foundation of socio-ecological models of health. Thus, my secondary findings in Chapter 4 – that the associations between the neighbourhood fast-food and formal PA environments and adiposity varied by household income – were largely confirmatory. In Chapter 8 I took fuller advantage of some of the longitudinal data available in UK Biobank, with the aim of building a stronger case for a tentative causal interpretation of those results. There, I expanded my focus to include more distal health outcomes (CVD- and cancer-related hospital admissions) and an additional exposure in the neighbourhood environment (greenspace). Building on the income-stratified analysis in Chapter 4, I assessed effect modification of these relationships by household income and area deprivation. I found that weak or null findings in the sample as a whole appear to obscure potentially important effects in some population subgroups, particularly people living in more deprived areas. Again, these results for effect modification by deprivation and income add to the rich, existing evidence base on the important role socioeconomic circumstances play in determining health.

Perhaps more importantly however, the results of Chapter 8 suggest that examining effect modification by multiple socioeconomic indicators in parallel and in combination, rather than in isolation, can yield greater insight into the potentially different ways in which various aspects of the socioeconomic circumstances of people's lives influence their relationship with the built environment and its effects on their health. Area-level socioeconomic characteristics are known to have an independent effect on health, over and above individual-level socioeconomic position[23], providing a rationale for considering both. The results from Chapter 8 reinforce this point and remind us that using area deprivation as a proxy for individual socioeconomic status will not always be appropriate. This may be especially true when assessing the health effects of specific features of the neighbourhood built environment. Personal socioeconomic resources will likely moderate those effects differently from how neighbourhood socioeconomic conditions will. Furthermore, the two in combination may be particularly important for determining an individual's sensitivity to the health-promoting and health-damaging characteristics of their neighbourhood built environment.

9.3.1.4. Context matters, and does so at various scales

Chapter 7 highlights the importance of considering potential geographical heterogeneity in any relationship between built environment and health – i.e. *where* do neighbourhoods matter? – and the implications for generalisability (or possible lack thereof) of findings from single-site studies if they are not carefully contextualised. These findings are consistent with similar evidence that has begun emerging from other recent studies – studies in a range of settings, that use various techniques to examine geographical heterogeneity in various neighbourhood health relationships[24–26]. For example, a study in the US showed substantial regional variation in the direction and magnitude of the relationship between an established measure of the neighbourhood food environment and BMI[26]. Another study in France examined geographical heterogeneity in the relationship between summary measures of the neighbourhood environment (derived from principal component analysis of a mix of perceived and objective characteristics) and active commuting, and again found variation in both the magnitude and direction of the estimated effect[27]. Whilst probably best interpreted as preliminary, such evidence of geographical heterogeneity nonetheless demands an explanation. Just as the role of contextual determinants of health is highlighted by the failure of individual-level socio-demographic factors to fully explain geographical variation in health[28] and health behaviours[29], similarly the persistence of variation across studies of associations *between* the neighbourhood built environment and health (after accounting for other neighbourhood-level and individual characteristics), suggests there are additional,

important contextual factors being overlooked. The "black box of places" described by Macintyre and colleagues[28] in 2002 remains far from fully elucidated almost two decades on, and the presence of geographical heterogeneity in neighbourhood-health relationships further underlines how much we still don't understand.

Here I argue that examining contextual effect modifiers, operating at neighbourhood scale and beyond, may offer additional insights. In Chapters 6 and 7 I examined greenspace in the neighbourhood and wider local authority as one such possible modifier of the influence of formal PA facilities on adiposity. I observed that, as hypothesised, neighbourhood availability of formal physical activity facilities was more strongly associated with adiposity in settings that lacked informal, green resources that might encourage and facilitate physical activity, and less strongly associated when informal, green resources were more abundant. It is not possible to determine whether these are causal interactions between formal PA resources and 'green' or 'natural' ones, but these results provide some support for that plausible explanation. Frohlich and colleagues[30] argue that epidemiological approaches are inherently limited, both methodologically and epistemologically, in their ability to explain how the contexts of people's lives influence health outcomes. While there is undoubtedly truth to this assertion, findings such as those I report in Chapters 6 and 7, as well as in Chapter 8 with respect to area deprivation, are examples of ways that epidemiological methods can bring to light potentially important interactions between contextual factors, and point to complex relationships that may warrant further investigation using both epidemiological and non-epidemiological methods.

The findings from Chapters 6, 7 and 8 support the idea that context matters for understanding relationships between specific neighbourhood characteristics and health, a notion described by Myers et al as the "context of context"[31]. The findings presented in Chapter 7, also suggest that the contexts in which neighbourhood-health relationships play out matter *at various scales*. This is one important element of the 'relational' perspective on understanding how place influences health advocated by Cummins and colleagues[11] in a seminal paper in the field, yet to the best of my knowledge it is rarely considered in the empirical epidemiological literature. The results of Chapter 7 indicate that not only do we need to consider the multi-dimensionality of the neighbourhood environment itself, as demonstrated in Chapter 6 and in studies in other settings (e.g.[32,33]); we should also consider the wider macro environment, i.e. the physical and socio-cultural contexts in which neighbourhoods are located (and arguably also the political and commercial contexts[34]). At the very least, we fall short if we stop at the point where we

identify an average association between some characteristic of the built environment and obesity, and infer from there that if we intervene on that characteristic we might expect to see improvements to population obesity risk. If these relationships vary over geographical space, influenced by the other features in a person's neighbourhood or by the wider context(s) in which that neighbourhood is located, such as a city or nation, then we cannot assume that interventions will be effective in all places. A more nuanced understanding of how, where and when the built environment influences obesity is needed in order to intervene effectively on it.

9.3.2. Understanding effect modification to guide public health intervention

Resources are always finite, and intervening in one place or in one way inevitably entails opportunity costs, so it is important to identify the settings and population subgroups in which a given intervention is likely to have the greatest impact. An examination of effect modification (on the additive scale) can inform us about this[35], and this is the approach I have taken throughout the thesis. Ignoring the modifying roles of other contextual features, whether they be other neighbourhood characteristics (as in Chapter 6) or attributes of the wider geographical area (Chapter 7), could lead to ineffective interventions and a waste of resources. It may also lead to 'failed' interventions when these are evaluated, and thereby undermine broader efforts to highlight and tackle contextual/socio-ecological/upstream determinants of health.

9.3.2.1. Understanding SES as a modifier points to ways to reduce health inequalities and avoid intervention-generated inequalities

In Chapter 8, I found that area deprivation and household income modified, on the (public health-relevant) additive scale, the association between availability of formal PA facilities and CVD-related outcomes, but did so in opposite directions. The results implied that greater availability of formal PA facilities may be particularly beneficial in deprived areas, but only for the those who can afford to access them. In terms of intervention, this suggests the greatest public health benefit would come from locating more PA facilities in deprived areas and ensuring they are low cost or free to use for those in low-income households. This conclusion is supported by evidence from a natural experiment in deprived areas of the North West of England, which found that free access to council leisure facilities was linked to improved health outcomes[36].

Similarly, results from Chapter 4 also indicate that while an association with adiposity was observed for all levels of household income, greater availability of formal PA facilities appeared, unsurprisingly, to disproportionately benefit wealthier households. Planners or

policy makers tempted to promote physical activity by investing in or incentivising the opening of more formal PA facilities may inadvertently produce intervention-generated inequalities[37] if they fail to address socioeconomic barriers to access. To avoid widening health inequalities it is critical that built environment interventions that increase availability of facilities such a leisure centres, gyms, sports fields and swimming pools offer genuinely affordable access for those with the most constrained household budgets.

9.3.2.2.Tailoring interventions to the local context to maximise health outcomes

In Chapter 6, I found evidence that formal PA facilities may buffer against a lack of informal, green resources for physical activity in areas where the latter are scarce, but that the potential benefits of formal PA facilities in terms of obesity risk may be undermined by an unhealthy food environment close to home. These results suggest that locating formal PA facilities in places with fewer public green resources and reducing the prevalence of fast-food stores in areas with formal physical activity resources, may maximise the health benefits to be derived from these neighbourhood resources. To my knowledge, this is an aspect of neighbourhood-health relationships that has not previously been studied in this way in the UK, and only rarely outside it. The findings regarding an interaction between the PA and fast-food environments and the conclusions I've drawn from those, are consistent with a similar study in the United States[33], in which the authors concluded that combined changes to the food and PA environments would have stronger and more consistent effects on BMI than changes to only one dimension or the other.

In Chapter 7, evidence of geographical heterogeneity in the magnitude of associations between the neighbourhood availability of formal PA facilities and BMI, and fast-food proximity and BMI, suggests that these two (and presumably other) neighbourhood characteristics might matter more in some settings than in others. The variation amongst English local authorities suggests the possibility of effect modifying factors operating at a larger scale. I tested two of these, and in keeping with Chapter 6's results but on a different scale, I found some indication that the availability of green spaces, in this case 'natural' land cover as opposed to urban parks, might be one such modifier of the formal PA environment. Importantly, modification by attributes of the wider context could theoretically apply to any neighbourhood characteristic.

Where evaluation of neighbourhood built environment interventions or natural experiments demonstrate (or fail to demonstrate) effectiveness in terms of population health outcomes, such results should therefore be interpreted in the context of the wider setting of the neighbourhoods involved, and the features therein that may have supported the success of the intervention. In a recent review of environmental interventions to

promote physical activity, Panter and colleagues[38] found that authors of intervention studies often noted the importance of context (variously defined) and cited features of the specific setting of an intervention that contributed to their evaluation results. However, these 'contexts' appeared to typically be compositional aspects of the study setting (e.g. demographic makeup of the population), while far fewer studies were reported as reflecting on contextual moderators in the physical environment, for example. Another recent review of population health intervention research concluded that recognition of 'context' was usually superficial and that intervention research would "immediately benefit from a more systematic and serious treatment of context"[39].

While the role of context may be recognised *post hoc* by evaluators, explicit examination of cross-level interactions between neighbourhoods and their wider context (geographical or otherwise) using both observational and experimental data offers the promise of a more robust understanding of *where* neighbourhood interventions might be most effective. This points to an avenue for further research that might identify important modifiers of built environment-health relationships at city or regional scales. Such new knowledge of contextual moderation could then inform future evaluations, and guide the targeting and prioritisation of interventions and local planning decisions, in order to maximise health in any specific setting.

9.4. Limitations

Each research paper in the thesis has its own limitations, and these have been discussed in the respective chapters. In this section I will discuss more broadly the sources of potential bias and error in the project overall. I then specifically discuss some of the challenges to causal inference faced by studies of this kind, what I have done to address them, and the implications for the findings of this thesis.

9.4.1. Selection bias

As with any cohort study relying on volunteers to participate, UK Biobank may suffer from selection bias, and this risk may have been heightened by the response rate of only 5.5%. Indeed the sample does show some evidence of 'healthy volunteer' bias[40]. Additionally, a substantial number of participants were missing data on key covariates, resulting in them being excluded from my analyses. This too may have caused selection bias. UK Biobank participants are slightly less ethnically diverse than the general UK population, and more socioeconomically advantaged, but non-representativeness is not unusual in large cohort studies, and is not necessarily undesirable[41]. However, unintentional non-representativeness such as in UK Biobank may increase the risk of selection bias[42]. In any

study – whether representative or not[43] – if the probability of participation in the study (i.e. selection) is influenced by either the outcome or the exposure *and* either a mediator or an unmeasured (or imperfectly observed) confounder of the exposure-outcome association, this will result in collider bias[44]. How large a bias and in which direction is difficult to gauge, although it has been suggested that if the exposure is not associated with selection, then collider bias is unlikely[43]. However, especially with cross-sectional analyses of baseline cohort data, it does pose at least a theoretical risk of spurious associations being detected[43,44]. For this reason it is important, just as it always is, that attempts are made to replicate novel findings and that results are interpreted cautiously and in the context of the wider evidence base, and triangulated with other forms of evidence[45]. Importantly given my focus on effect heterogeneity, the large sample size of UK Biobank ensured sufficiently large subgroups across various third variables that represented potential effect modifiers. As Rothman and colleagues[46] have pointed out, balanced numbers of study participants across levels of a hypothesised modifying variable is more efficient than overall representativeness when it comes to assessing effect heterogeneity.

9.4.2. Information bias/measurement error

9.4.2.1. Measurement error and misclassification in environmental measures

The reliance in this thesis on the environmental measures available in UKBUMP, and to a lesser extent, the additional off-the-shelf greenspace measures from the GLUD, introduce some concerns about measurement error and the risk of misclassification bias. Unfortunately, the source documentation associated with UKBUMP provides only limited information about the dates and other details of primary data collection, and the accuracy and validity of the underlying databases cannot readily be assessed. Of particular concern is the risk of measurement error in the food environment measure. The food outlet classification in the source database is supplied by local authorities and may include misclassification of some outlets as restaurants rather than fast-food outlets, potentially overstating the true distance to the nearest fast-food outlet. If the quality of the environmental data varied geographically, as is plausible given its origins in local authority records, this could have magnified errors.

The neighbourhood exposure measures may also suffer from misclassification arising from the timing of their measurement. The UKBUMP source documentation cites the 2012 version of the source database (OS AddressBase Premium), but is somewhat ambiguous about the precise timing of the recording of the relevant land use types in that database. There is therefore a risk that the measured exposures may not reflect true baseline

neighbourhood exposure, especially for those recruited in the early stages of the study, if there have been changes to the neighbourhood environment in that time.

If any measurement error in the environmental exposures was random, it would be expected to attenuate observed associations[47], meaning the weak associations detected for proximity to fast-food outlets and adiposity (Chapters 4 and 5), and in Chapter 8 for associations with CVD and cancer, may be conservative estimates of the true effect. As Smith and colleagues[48] have pointed out in the context of their use of UKBUMP, by categorising the exposure measures (as I have) some of the risk of measurement error may have been mitigated.

Measurement error in any of the effect modifiers may have led to misidentification of the presence or absence of effect modification. Genetic risk, sex, and area deprivation should all be highly accurate. Fast-food proximity in Chapter 6 may be misclassified as discussed earlier. In Chapter 7, allocation of individuals to the wrong local authority due to the reliance on imprecise address coordinates may have introduced random error into the analysis of the two macro-environmental variables as potential effect modifiers.

With respect to the measure of neighbourhood greenspace used in Chapter 8, this measure may not adequately capture the full extent of the relevant greenness of one's neighbourhood, because the GLUD classification scheme does not include smaller parcels of greenness such a street trees. Some alternative measures of greenspace, such as NDVI, are more sensitive to overall greenness, and these have been analysed with respect to adiposity in another UK Biobank study[49], but this measure is only available for participants recruited to a subset of the UK Biobank assessment centres. As the GLUD measure is based on 2005 data, it may more accurately reflect the true exposure to neighbourhood greenspace for those participants recruited earlier in the baseline phase than those recruited later.

9.4.2.2. Spatial misclassification

Earlier in the thesis (Chapter 3) I raised the issue of spatial misclassification in its various forms, including the 'residential trap'[50] and the Uncertain Geographic Context Problem[51]. With respect to conclusions about the assessment of effect modification, for a given form of spatial misclassification of a health-relevant exposure to be a meaningful problem, it would need to qualitatively impede assessment of whether the primary exposure-outcome association varies according to values of some third variable. This might arise if all subgroup estimates were biased toward the null, leading us to make a Type 2 error i.e. erroneously conclude there was no effect modification when there actually was. It could also arise if misclassification systematically biases the estimated effects of the exposure

more in one category of a potential effect modifier than in another. What would this entail? If we think about one of the primary exposures in this thesis, the availability of formal PA facilities, defined as the number within a 1-km buffer around the home, then in Chapter 5 it would mean that at some levels of genetic risk of obesity, a 1-km neighbourhood was a more accurate representation of a causally-relevant geographical context than it was at other levels of genetic risk. It is not obvious that there is any *a priori* reason to expect this to be true. In Chapter 6 it would mean that a 1-km buffer would have to be a less appropriate delineation of neighbourhood for people in one category of fast-food proximity than those in another category, or in one category of park access than another. Again, it is not obvious that this would be the case. We may, however, think that the boundaries of a health-relevant neighbourhood might be different for some population groups. An obvious example is age: the size of the health-relevant neighbourhood area might be different for older people compared with younger people. It may also vary along socioeconomic lines, if for example car ownership, employment status, or social capital expand the area of one's causally relevant physical environment. If the latter is true, it is possible that the analyses of effect modification by household income and area deprivation in Chapters 4 and 8 may be biased, although it is not necessarily clear whether subgroup differences would be artificially widened or narrowed as a consequence.

9.4.3. Confounding

The breadth of the UK Biobank resource allowed for comprehensive adjustment for possible confounding variables, but like all observational studies, there is always a risk of residual confounding, either through omitted confounder variables or misclassification of measured confounders. Chapter 4 included a negative control analysis (using height as a control for adiposity) to check whether the estimated main associations between either the food or physical activity environment and adiposity were likely to be residually confounded. The results for the availability of PA facilities indicated possible residual confounding. Comparison of standardised coefficients across the relevant models suggests this residual confounding would only partially account for the observed effect, but that leaves the possibility that the estimates for that exposure are overestimated. Interestingly, the food environment models did not appear residually confounded. It may be that height is an imperfect control for adiposity, with respect to either or both the exposures examined. Adult height should be independent of neighbourhood environment in mid-life, but there is a possibility that adult neighbourhood is correlated with childhood determinants of height (e.g. inadequate nutrition, adverse events). Height is negatively

correlated with low birth weight in this sample, but low birth weight doesn't show substantial bivariate correlation with either the physical activity or fast-food environment.

I can identify four main potential sources of residual confounding in this thesis. First, a commonly identified problem for neighbourhood effects studies is the possibility for self-selection of healthier individuals into more health-promoting neighbourhoods in order to meet their lifestyle preferences or needs. If factors that drive selection into a particular neighbourhood type also predict the outcome, this may introduce confounding, and this is not something that I was able to adjust for in the analyses presented here. I expect that at least some of this risk will be mitigated by the residential stability of the UK Biobank cohort (>60% of the sample have lived at their current address for at least 10 years, with a mean of 18 years). Furthermore, for those with substantial choice over where they live, the presence of PA facilities alone is unlikely to drive that choice, and empirically there is not a strong correlation between household income and neighbourhood availability of PA facilities in this cohort. That said, the negative control analyses showing some residual confounding of associations with the PA environment but not of fast-food environment might indicate that physical activity preferences are a stronger driver of residential self-selection than food preferences are. Studies that have directly examined the influence of self-selection on neighbourhood-health effects have reached inconsistent conclusions about the likely bias this may induce[52–54], and most have been focused on transport-related physical activity and in other settings, so it is not clear how large a threat this is.

Second, a related source of confounding, referred to elsewhere as indirect self-selection[55] may arise if 'deprivation amplification' is operating in the UK. Deprivation amplification is the idea that in more economically deprived areas the availability of health-promoting neighbourhood resources is lower[56]. But with respect to the distribution of PA facilities across social gradients, previous research in the UK has revealed a mixed picture[57,58]. I adjusted models for area deprivation and three measures of individual socioeconomic status (income, education and employment) in an attempt to overcome this.

Third, I observed in Chapter 4 that where intermediate models were not adjusted for individual socioeconomic characteristics, coefficients were biased away from the null. If in the final, fully adjusted models I have not been able to adequately control for individual socioeconomic position (e.g. because of lack of specificity in the categorical, non-equivalised income variable), the main effects may be slightly overestimated.

Fourth, and of particular concern with respect to the fast-food exposure measure, are the possible confounding effects of unobserved characteristics of the neighbourhood environment. While absolute measures of neighbourhood access to fast-food outlets are

common in the literature, it is increasingly being recognised that the relative opportunity to access healthy or unhealthy food options may be of greater consequence for diet and health[59]. Several studies have found that simultaneously accounting for healthy and unhealthy food outlets in an area yields larger and more precise estimates of health effects than when considering only a single dimension of the food environment[19,60,61]. It was not possible to account for other dimensions of the food environment using UKBUMP, and this means the regression coefficients may be biased – probably towards the null because healthy and unhealthy food stores tend to cluster together.

In terms of confounding, there is also a possibility that if any of the main associations are confounded in one stratum of the potential effect modifier and not another, I may have erroneously inferred effect modification when none is present. Alternative negative control analyses might shed more light on how much of a risk residual confounding truly is in these (and other) analyses, although suitable controls are difficult to find.

9.4.4. Reverse causation

The possibility of reverse causation in the cross-sectional analyses in this thesis (Chapters 4-7) must also be considered, because like any cross-sectional studies there may be uncertainty with respect to the temporal sequence of the exposure and the outcomes. Movement of people between neighbourhoods over time may give rise to reverse causation in cross-sectional analyses if, for example, individuals with lower adiposity choose to live in areas with more PA facilities. Current adiposity may also reflect exposure to neighbourhood environments earlier in life, posing a further challenge for establishing the necessary temporal ordering of the exposure before the outcome. Furthermore, as mentioned in Chapter 4 the relationship between commercial features in a neighbourhood (including takeaways and gyms) and health is likely to be bi-directional, if retailers target their businesses at areas of higher demand, meaning the estimated magnitude of an association may overstate the effect of the neighbourhood characteristic on the health outcome. It is likely that both supply of and demand for neighbourhood-based food and PA resources drives the use of these resources. Interestingly, evidence from the US suggests that supply of fast-food outlets is a greater driver of increased fast-food consumption than demand for them is[62]. Nonetheless, demand-driven location of neighbourhood resources is also likely to some extent – with retailers selecting sites for new fast-food outlets of PA facilities in places where they expect demand to be high. But this also presupposes that availability of neighbourhood resources will influence behaviours such as diet and physical activity, reinforcing the idea that the relationships being examined in this thesis are to some extent bi-directional and self-reinforcing. These

issues highlight the need for longitudinal studies that capture changes in the neighbourhood environment as well as changes in health outcomes, and the potential value of taking a systems approach to modelling these relationships[63].

Whether these issues might be more or less pertinent for a residentially stable population such as the UK Biobank cohort, is very difficult to assess. Almost certainly any information available to retailers about a local population is more likely to be accurate for a residentially stable population. However, it does not necessarily follow that a residentially stable population is more (or less) likely to be a target for new resources. Arguably, a less stable population (younger, less established in place) may be more appealing to businesses establishing a new fast-food outlet or new sports facility, because they may be more likely to change their behaviour in response to new stimuli. Further complicating the picture, the main explanation for the residential stability of the UK Biobank cohort is likely to be its relatively older age[64,65]. The cohort is geographically distributed across a wide range of neighbourhood types, and (whilst somewhat more socioeconomically advantaged than the target population) spans the breadth of middle-aged UK society. As such, participants in the study are embedded within a range of mixed-age neighbourhood populations, and while they may themselves tend to be long-term residents of their area due to their age (and so be less likely to have been exposed to other causally relevant residential neighbourhoods in recent years), they may not be representative of the residential stability of their local community more broadly, and therefore of the likelihood that their area is targeted by retailers based on local demand for particular resources.

Within this project it has not been possible to examine these risks, but the theory behind neighbourhood effects posits that neighbourhood causing adiposity is likely to be the dominant pathway.

9.4.5. Distal vs proximal outcomes

In this thesis I have chosen to focus on outcomes that are relatively distal from the exposure in terms of the assumed underlying causal process. There are several reasons for this. The first is that I wanted to look at exposures on both sides of the energy balance equation (and in the case of greenspace in Chapter 8, exposures that might influence other pathways to health). This is an attempt to recognise the complexity of both chronic disease and of neighbourhood environments. Rather than attempting to isolate single, linear pathways from one exposure to a health behaviour (already the aim of the majority of studies), I was more interested in how factors across multiple levels might work in concert to influence complex outcomes such as obesity. Second, the behavioural data (dietary intake, physical activity) in UK Biobank that are available for the full sample are

subjectively measured, while the outcomes I have used are objectively measured adiposity, and linked administrative health records. The IPAQ short form (the measure of physical activity available) has been shown to substantially overestimate physical activity compared to objective measures[24], and the dietary data are very incomplete and recorded only for a biased subsample, as I showed in Chapter 4. Nonetheless, by focussing on distal outcomes where the presumed causal relationships are mediated by behavioural and cognitive processes, inference and interpretation can be more problematic, and the risk of random error is greater, making it harder to detect effects.

9.5. Implications for causal inference

Like the bulk of this PhD research, most research on the influence of neighbourhood environments on health is observational and cross-sectional[66], limiting the ability of researchers to draw causal inferences about neighbourhood effects. Randomised trials are very difficult to implement for the study of neighbourhood effects, so we usually have little choice but to rely on observational data. While as good researchers we always include the mandatory caveats about not inferring causality from observational data, and the estimates being associational, the often unspoken aim of these studies remains to estimate something approximating a causal effect, albeit while acknowledging the challenges to doing so, and the unavoidable risk of confounding. Curtailing our language to avoid causal claims, we nonetheless tend to implicitly interpret results, at least tentatively, as if they point to a causal effect[67]. The same is true in this thesis: within the limits of the data available, and the inevitable risk of confounding, I have sought to minimise bias in the estimates produced, so that the direction (if not the precise magnitude) of the associations might be cautiously interpreted as reflecting an underlying causal effect.

9.5.1. Challenges to the main assumptions required for causal inference

Causal inference rests on three main assumptions: exchangeability (no unmeasured confounding); positivity (sufficient overlap between exposed and unexposed within strata of all covariates); and consistency (the exposure is sufficiently specific that different variants of it cannot have different effects on the outcome)[68].

Observational studies are especially at risk of *exchangeability* violations, because of the risk that important confounders are not fully observed. I summarised the likely sources of residual confounding in the preceding pages. I sought to overcome this by using DAGs – informed by background knowledge – to help identify likely confounders and thereby inform my covariate selection for regression modelling. The negative control analysis in Chapter 4 was designed to identify whether residual confounding was likely to remain. In

each chapter, where there was doubt about adequate confounder control, I attempted to include a considered discussion of the possible impact on the results.

The *positivity* assumption has been identified as particularly challenging for neighbourhood effects studies, with violations potentially arising when residential segregation occurs along socioeconomic and racial lines[69] or where there is strong self-selection (as described above) into areas. This can lead to the dataset containing 'exposed' individuals for whom there is no equivalent 'unexposed' individual, and models relying on extrapolation and "off-support" inference[70]. The problem is sometimes referred to as structural confounding, and it can only be partially mitigated by model adjustment for individual predictors of residence[69,70]. Fortunately, the very large sample size of UK Biobank should mitigate against severe non-positivity. Where I used categorical variables with more than a few groups (e.g. ethnicity), I undertook preliminary analyses to ensure model estimates were not sensitive to the number of categories, as this might have indicated sparsity issues that could lead to non-positivity.

Finally, the *consistency* assumption requires that an exposure is defined with sufficient specificity that different variants of it do not have different effects on the outcome[68]. 'Bundled' or 'compound' exposures – such as neighbourhood typologies – can violate this assumption because changes to components of the exposure may have different effects on the outcome[68,71]. In Chapter 6 I avoid this problem by considering how the estimated effect on adiposity of one neighbourhood characteristic is modified by another neighbourhood characteristic, rather than how a bundle of neighbourhood characteristics influences adiposity. But other kinds of neighbourhood measures used routinely may also violate the consistency assumption. To use an example from this thesis, having two formal PA facilities in a one-kilometre neighbourhood buffer might equate to having one gym and one swimming pool for one person, but two public football fields for someone else. In a strictly causal sense, it is not then clear what it means to say that having two PA facilities in the neighbourhood causes a lower BMI than having none. Facilities of differing quality, affordability or capacity raise similar issues. Interpreting results using these measures for policy, it's not clear what intervention the findings imply would have an impact on population BMI. This is not something I've been able to address, due to my reliance on secondary data, but as discussed earlier, the lack of precision in measures used in this thesis is likely to have resulted in non-differential misclassification, leading to conservative point estimates of the main associations[47]. If such misclassification is also non-differential across levels of effect modifiers, I may also have underestimated the magnitude of differences between those subgroups.

9.5.2. Causal effects or close enough?

New studies that may be better placed to estimate true causal effects of neighbourhoods are emerging: as causal inference methods develop; as large, population-based panel surveys mature and provide rich longitudinal data; and as quasi-experimental study designs are embraced by social epidemiologists. So far, a number of these more causally focussed studies suggest no truly causal effects of neighbourhood over and above the composition of the populations living there (e.g.[72,73]). However, such studies often impose far stricter inclusion criteria than studies such as mine, for example restricting analysis to 'movers' – people who relocate between neighbourhoods during the course of a longitudinal study. While this approach has appeal, it is also the case that movers are likely to be different from non-movers in important ways. If neighbourhood effects are heterogeneous across population subgroups, and movers are less likely to belong to those groups that are more sensitive to the built environment, effects will be underestimated if we look only at movers. Or rather, findings will not be generalisable to non-movers, who comprise a far larger segment of the population.

Whether we can ever really estimate anything resembling a causal effect in neighbourhood studies using observational data remains unclear[74]. Certainly this thesis suggests there is no such thing as a single effect of a given neighbourhood characteristic – rather there are probably many, depending on the distribution of effect modifiers in the population. But through triangulation[75], inference to best explanation[76], and transparency about the assumptions and likely bias in our studies[77], we might at least settle upon 'good enough' evidence on which to base highly plausible working hypotheses about the causes of complex health conditions[78]. These can then be used judiciously to guide public policy and urban planning decisions, the real-world impacts of which can then be evaluated against our hypotheses, and our hypotheses updated accordingly. This process was described by the Foresight report as a "virtuous circle"[9] and recognised as necessary given the pressing need to tackle obesity.

9.6. Overall strengths

The strengths of this project fall into three categories: strengths of the data used, novelty, and methodological rigour.

9.6.1. Strengths of UK Biobank and the UKBUMP

In this project, I have made use of a unique cohort that is large and heterogeneous enough to explore multiple effect modification hypotheses. Feasibility constraints on large-scale studies of the built environment are likely to be an important reason for the relative

scarcity of effect modification analyses or explorations of geographical heterogeneity. With upwards of 300,000 individuals available with complete data for each analysis in this project, UK Biobank and the UKBUMP provided the opportunity to work with a sample sufficiently large to draw reliable conclusions.

Setting aside the limitations of the UKBUMP data described earlier, the UKBUMP nonetheless represents a unique and valuable resource for the investigation of the health influences of neighbourhood environments. It has provided the opportunity in this thesis to simultaneously examine multiple neighbourhood exposures that were objectively derived from detailed, routinely collected national data, and ascertained for each individual on the basis of their precise home address. Person-centred residential environment data of this kind for such a large sample is unprecedented in the UK. In the first research paper of the thesis (Chapter 4), I therefore started by taking advantage of this dataset to replicate studies done on smaller scales and in other settings.

UK Biobank has deliberately sampled a population at an important stage of the lifecourse: mid-life is a critical time for the development of chronic disease, and it is where the burden of obesity, cardiometabolic disease, and many cancers – and their associated healthcare and wider economic costs – are concentrated.

The wealth of objectively ascertained outcome measures in UK Biobank strengthens the internal validity of the study, by minimising measurement error. In Chapters 4-7 of this thesis I used measures of adiposity collected by trained staff using standardised techniques, including data collection with bioelectrical impedance machines that transferred weight and body fat measurements directly to the assessment database, eliminating the risk of human error in the data entry phase. The consistency of findings across multiple objective measures of adiposity lends additional validity. In Chapter 8 I used official NHS records of hospital admissions, again minimising the risk of random or non-random error in ascertainment of the outcomes.

The breadth of the UK Biobank resource, spanning clinical outcomes and genotypes, health behaviours, anthropometry and detailed socio-demographic data, combined with multiple environmental measures and the opportunity to link additional spatial data, also ensured I was able to comprehensively adjust regression models for a wide range of potential sources of confounding.

9.6.2. Novelty

Despite an increase in research considering potentially differential neighbourhood effects according to sociodemographic factors, there has been limited investigation of other

potentially important modifiers, particularly across multiple levels of influence. Some of the evidence gaps exist because of limited availability of appropriate data. While UK Biobank is subject to its own weaknesses, it does, with its size and scope, present a unique opportunity to examine sources of effect heterogeneity that other studies have been unable to consider empirically. In this thesis, I have made use of the UK Biobank resource to contribute novel evidence addressing some of these evidence gaps.

In Chapter 5, I presented novel evidence suggesting that genetic risk of obesity may modify the response of individuals to their local food environment. While these findings need to be confirmed by other studies, they contribute to a small, emerging body of evidence for GxE interactions in relation to obesity where the 'E' is a truly environmental variable. Understanding GxE interactions of this kind may have implications for understanding the aetiology of obesity and the biological pathways by which obesogenic environments get 'under the skin'. Chapter 5 can therefore be seen as contributing to what Galea & Link[79(p.4)] describe as important research at "the intersection of factors that matter inside and outside the skin".

Chapter 6 is a surprisingly rare example of an investigation of interactions between multiple dimensions of the neighbourhood environment. Despite being grounded in a theoretical perspective that recognises and indeed brings to the fore the role of local contexts as part of a complex set of obesity determinants, research on relationships between the built environment and obesity has largely neglected to explicitly examine the way the various dimensions of neighbourhood environments operate in concert. The exceptions are studies that look at composite measures of neighbourhood obesogenicity, or similar constructs such as neighbourhood typologies. However, such measures pose challenges for the consistency assumption on which causal inference relies[71]; in terms of both aetiology and identifying the contexts where intervention might be most effective, they lend themselves to more general conclusions about the importance of holistic urban planning for health. Here I have sought to take a more explicit approach to a similar question, and show specifically that the association between availability of formal PA facilities and adiposity is somewhat stronger among people living in places lacking parks and other open green spaces, and weaker among people living closest to a fast-food store. Both findings lend support to the hypotheses I set out to test in Chapter 6.

In Chapter 7 I explored, for the first time across a large area of the UK rather than within a single city, whether neighbourhood-health associations vary geographically. Having established that such geographical heterogeneity does seem to exist for the two

relationships I used as examples, I then demonstrate one approach to exploring possible contextual drivers of geographical variation.

Each of these novel investigations adds detail and complexity to existing theoretical models of how neighbourhoods influence health, and the findings establish a rationale for exploring these relationships further.

9.6.3. Methodological rigour

Most of the evidence I present in this thesis is based on careful cross-sectional analysis of observational data, but it is not strictly possible to draw inferences about causality. I have, however, carefully selected and applied appropriate analytical strategies that included a rigorous approach to identifying and mitigating likely sources of bias. I examined these possible biases with sensitivity analyses, and reflected on them when discussing the results, so that the findings I report serve as a solid foundation on which future studies might be based using data that supports more causally focussed study designs. I discuss some of these possibilities later in this chapter, and I covered some of the issues around causal inference earlier in this chapter.

In the research paper in Chapter 4, the main associations between each of the fast-food and PA environment exposures and adiposity were estimated after controlling for the influence of the other environmental exposure, in recognition of the fact that the two may be correlated and share a common causal antecedent, as well as both potentially influencing the outcome[80,81]. Importantly, this adjustment, along with adjustment for residential density resulted in the estimated associations being larger in magnitude, suggesting that failure to adjust for confounding by other environmental variables may attenuate model estimates. Many studies of these relationships do not adjust for potential confounding by other factors in the local environment, so this is an important secondary finding of the study.

While we must be cautious about overinterpreting subgroup/interaction analyses from a single sample, especially using cross-sectional data, these analyses were based on *a priori* hypotheses about plausible modifying relationships, and the conclusions are internally consistent within the thesis. At the very least, they provide preliminary support for the idea that relationships between the food and physical activity environments and obesity-related outcomes are very likely to be modified by numerous influences at individual and environmental levels. These findings are not surprising, but they provide an empirical base on which to build more evidence to inform tailored built environment interventions.

9.7. Policy implications

Translation of research evidence into policy relies on assumptions of generalisability. Such assumptions can be undermined by a failure to recognise heterogeneity of effect across population subgroups and geographical space, and the likely drivers of that heterogeneity. Policy responses based on studies ignoring effect heterogeneity may fail if average estimated effects conceal important information about where and for whom an intervention is likely to work[82,83]. Ignoring effect heterogeneity also opens up the risk of widening health inequalities, wasting resources (e.g. by concentrating resources in places that need them least or where competing influences in the built environment will dilute their potential impact) or intervening in places where the need is smallest. Dahlgren & Whitehead[84(p.22)] argue that "a comprehensive health strategy… [should] include both downstream and upstream determinants of health and the relationships between the two, as they are often interlinked closely". Some of the nuance of the relationships between upstream and downstream determinants of health has been elucidated in this thesis, and emphasises the importance of the interplay between health determinants at multiple levels of influence. The importance of taking context into account when designing interventions and intervention research has been highlighted elsewhere[85,86], and this thesis reinforces that point.

This thesis provides cross-sectional evidence that the influence of particular characteristics of residential neighbourhood environments varies across a range of factors at multiple levels, and therefore cannot be assumed to be uniform. In terms of policy and interventions, there are two ways to respond to evidence that the influence of the built environment is not uniform across the population:

1. Ensure equal allocation of healthy resources so that everyone has access to the same opportunities to engage in healthy behaviours.
2. Allocate healthy resources differentially, to support people to overcome unequal risks arising from other factors (an equity approach).

The results of the GxE analysis in Chapter 5 give support to Approach (1) - making environments healthy for everyone so those at greater genetic risk are not further disadvantaged. On the other hand, the results of Chapters 4 and 8 point to some practical ways that Approach (2) could be applied, e.g. improving the physical activity environment in deprived areas to maximise population health benefits, and ensuring affordable access to recreation facilities. The results of Chapters 6 and 7 show either approach should also take into account physical attributes of the local and wider context.

Some policy implications relating to specific exposures are discussed below.

9.7.1. Formal physical activity environment

As detailed earlier in the synthesis of key findings, several findings relating to the availability of PA facilities point towards some specific policy recommendations. In all five research papers, a clear inverse association was seen with either adiposity, CVD or cancer, in at least some places or subgroups. This suggests that increasing availability of local PA facilities may be an effective strategy for improving population health. Critically though, those health benefits appear unlikely to accrue uniformly across the UK or its mid-aged adult population. The results in Chapter 8 suggest that the greatest public health benefit would come from locating more PA facilities in deprived areas, but only if access was affordable for those on low incomes. Otherwise, failure to ensure equity of access could widen health inequalities. The findings in Chapter 4, that PA environment effects appear to be stronger among those with higher incomes, also imply that PA facilities need to be accessible to all. This could be achieved through subsidised access fees and incentives for low-cost operators to open in deprived neighbourhoods. Greater central funding to local authorities for leisure and recreation services, tied to a requirement to provide affordable access, may be another approach. Research has shown that free access to council leisure facilities can produce improved health outcomes in deprived areas of the UK[36].

The findings from Chapter 6 suggest that locating formal PA facilities in places with fewer public green resources and reducing the prevalence of fast-food stores in areas with formal PA resources, may maximise the health benefits to be derived from these neighbourhood resources. Further research on macro-environmental factors that might also be important modifiers may provide additional policy guidance to local authorities and regional planning authorities, to help target interventions.

Finally, the fact that no GxE interaction was observed involving the formal PA environment suggests that, other modifiers aside, we may all (irrespective of genetic risk) be equally positioned to benefit from making our neighbourhoods places where it is easier to be physically active.

Summary of recommendations for the formal physical activity environment:

- Increase availability of affordable formal physical activity facilities in deprived areas.
- Consider wider context and other dimensions of the neighbourhood when intervening on the residential built environment to improve health.

9.7.2. Fast-food environment

The findings from Chapter 5 suggest that reducing the proximity of fast-food outlets to residential areas may be more beneficial for those at increased genetic risk of obesity. That some of us may, due simply to our genetic make-up, be more sensitive than others to elements of the local food environment, only serves to underline the importance of taking steps to ensure our neighbourhood environments promote, rather than inhibit, good health. These results do not imply that environmental interventions should target those at higher genetic risk – this is obviously not possible – but rather that we should recognise that variation in genetic risk may leave some of us more vulnerable than others to unhealthy environmental conditions. As we cannot control our genetic makeup, it is unreasonable to demand and expect that people at greater genetic risk work harder than others to overcome the influence of unhealthy food environments.

In the era of a move towards 'precision medicine', evidence of GxE interactions is sometimes framed as holding great promise for customising behaviour change interventions to match a person's individual genotype[87,88]. This runs counter to the goals of most who take a socio-ecological perspective on health. By focussing further upstream, on truly environmental (rather than behavioural) factors that might interact with genotype, a different framing of the usefulness of GxE interaction studies becomes possible. If people at increased genetic risk of, for example, obesity, are shown to be more sensitive to environmental cues, it may help explain why some people are "better" at taking "personal responsibility" in an obesogenic environment (to use the language of the dominant narrative). If genetic risk can be compounded by an unsupportive food environment, which may also interact with a lack of socioeconomic resources, that may well be a cocktail for poor health. GxE interactions may help us develop more complex models of obesity and other important chronic diseases. In the long run this might help shift the narrative away from one of blame and individual responsibility. Results such as these offer a strong argument for a population-wide approach: if those with greater susceptibility to high BMI are most sensitive to an unhealthy food environment, then population-level environmental interventions might be of most benefit to those at greatest genetic risk.

Summary of recommendations for the fast-food environment:

- Recognise the burden placed by unhealthy neighbourhood food environments on individuals who are at increased genetic risk of obesity, and limit the presence of unhealthy food outlets close to all residential areas.
- Move away from policies and rhetoric that focus on individual responsibility.

9.7.3. Neighbourhood greenspace

Neighbourhood greenspace was a secondary focus of the thesis. Consistent with the emerging evidence from many other studies, the findings from Chapter 8 suggest that greenspace near home may influence health outcomes other than cardiometabolic ones, though in this study the only evidence of an association appeared to be with cancer, among people in deprived areas. Other evidence also suggests greenspace is important for health in deprived areas[89-91]. My findings therefore lend further weight to calls for improving access to greenspace in deprived areas. Increasing availability of greenspace can also provide additional benefits and ecosystem services, such as providing habitat for other species, mitigating against pollution and flooding, and moderating temperature and noise[92]. Unfortunately, safety and aesthetic concerns mean the particular management strategies of public green spaces to maximise use by humans do not always align with the best management strategies for maximising other ecosystem services (e.g. biodiversity conservation)[93]. These tensions need to be managed with an eye to balancing priorities and maximising benefits across a range of important outcomes.

Summary of recommendations for neighbourhood greenspace:

- Improve access to quality greenspace in deprived areas.
- Manage greenspace in a way that maximises health benefits without compromising ecosystem health and services.

9.7.4. Global challenges for policy

The increasing urbanisation of our world is exposing more people every day to the various features of urban living that may influence our risk of chronic health conditions[94,95]. Understanding how the environments in which we live influence our health is critical to managing the growing burden of obesity and non-communicable disease. Most of the growth in urban areas is occurring in low- and middle-income countries, yet most of the research on the health effects of urban living environments is conducted in already highly urbanised high-income countries. On the one hand we might argue that by learning as much as we can from highly urbanised countries such as the UK, about the processes by which 'obesogenic' and other health-damaging features of our cities interact with other factors to cause ill health, we should hopefully be able to apply those lessons in rapidly urbanising areas around the world. But on the other hand, this thesis demonstrates that even within a single country research findings are likely to be context-dependent and not necessarily generalisable from one setting to another. There is therefore a need for a greater investment in this kind of research in a wider range of settings globally, including

the creation of appropriate data resources such a cohort studies with linked environmental measures. It may also be valuable to conduct cross-national studies across a wider range of settings, with the explicit aim of understanding more about when findings about the effects of built environment on health are generalisable to other settings and populations, when they need to be adapted for the local context and supplemented with new local data collection and knowledge generation, and ultimately, what that means for policy action.

Perhaps most importantly, the policy solutions to address complex problems like the high prevalence of obesity will inevitably be themselves complex and multidimensional[9,10]. Built environment interventions are likely to be a part of that, but will need to be accompanied by a coordinated suite of other changes across multiple policy domains and all levels of society, informed by multidisciplinary evidence and led by government.

9.8. Future research directions

Various research gaps and challenges remain in the field of neighbourhoods and health. The work of this PhD draws attention to some in particular, and also opens up new questions. In this section, I highlight some of these directions for future research, classifying them into two broad categories: (1) further explication of potentially heterogeneous neighbourhood effects (and drivers thereof) within a socio-spatial epidemiological framework; and (2) potential contributions from other disciplines and methodological traditions. I then identify some suggestions for specific additional studies using the UK Biobank resource, and specific extensions to UK Biobank that would enable future research in the field of neighbourhoods and health.

9.8.1. Directions for future research

9.8.1.1. Further socio-spatial epidemiological examination of heterogeneous neighbourhood effects

First, further studies attempting to refute the findings presented in this thesis are needed, utilising other existing and future cohorts. In light of the likely selection bias in the UK Biobank arising from the low response fraction, it is especially important that replication studies are done in more representative samples. It would also be valuable to see if some of the specific findings were replicable in a local setting, based on the expectations from this project (e.g. within a single city with detailed data, is the association between formal physical activity facilities and BMI modified by proximity to fast-food outlets, as I observed in Chapter 6 for the UK as a whole).

Further research to identify additional, contextual macro-environmental modifiers of neighbourhood effects is also warranted, especially given those I examined in Chapter 7

showed only weak evidence of interaction, implying there are other factors driving the geographical heterogeneity observed in that chapter. Consideration of other scales is also needed. It has been suggested that greater attention be paid to the political and economic contexts in which geographical inequalities in health are produced: for example, Bambra and colleagues[34] recently laid out a strong case for adopting a political economy approach to the study of place and health, in which the structural forces of politics and economics that operate at national and international scales are recognised as shaping the more traditional, and proximate, social determinants of health. In this vein, we might also expect these structural forces to influence how and where more local, neighbourhood level factors influence health outcomes.

There are also additional, related studies that would contribute to the objectives of this thesis but were not possible within the constraints of the available data in UK Biobank. For example, neighbourhood influences across the life course are likely to be important[96], but there were no data on these in UK Biobank.

As mentioned above, rapidly growing urban areas in low- and middle-income countries present opportunities to avoid making the same mistakes again, by applying existing and emerging knowledge to new cities. The generation of locally relevant, setting-specific evidence is important. But alongside this, there are rich extant data resources in high-income countries amenable to further interrogation and linkage to geospatial environmental data, providing immediate, low-cost opportunities for additional research. Bearing in mind the need to consider context and possibly heterogeneous effects, these resources should continue to be leveraged to deepen and strengthen the evidence base, in parallel with new data generation in understudied areas globally.

Population health intervention research that builds on the evidence base(s) to which this thesis makes a small contribution is much needed. Observational research, especially cross-sectional studies like Chapters 4-7 here, can suggest, but not clearly identify, which interventions or policies may have the biggest benefits for health. More studies are needed that explicitly test the effects of planning and land management policies and local built environment interventions. To this end, there is also scope to take advantage of natural experiments[97]. These can raise issues of generalisability posed by possible effect heterogeneity, as they tend to be context-specific or may only target particular population subgroups[82], so comparison of similar or related natural experiments in different contexts may provide richer insights in terms of contextual modifiers.

9.8.1.2. Contributions from other disciplines and methodological traditions

Contributions from other disciplines and methodological traditions would deepen our understanding of the social processes driving some of the effect modification relationships observed in this thesis. Sociology and anthropology, for example, offer theoretical and methodological frameworks that are invaluable for understanding the social meanings of health-related behaviours such as eating and exercising, and how these interact with physical, place-based resources. Qualitative, participatory and ethnographic methods can reveal aspects of people's lived experiences of their neighbourhood environments that are inaccessible to the quantitative researcher[98,99]. This has been demonstrated by various studies of how people engage with and navigate the food environment, and how these processes are informed or constrained by individuals' identities and socioeconomic resources[100-103]. As well as standing alone, such approaches help to make sense of quantitative findings, illuminating *why* neighbourhood characteristics matter for health, and may generate further hypotheses along the lines of those tested in this thesis[104].

Sociological, political and economic theory and methods may provide further useful perspectives on understanding the ways that various structural forces at the national and international level influence social and physical determinants of health at the neighbourhood and individual level[34], and how commercial and industrial forces take advantage of the social meanings we attach to health-related behaviours[105].

Systems thinking offers another way to approach this area of research[10,106]. Complex health outcomes such as obesity and NCDs, and the inequalities in their distributions, have been viewed (by some) in these terms for several years, with neighbourhoods recognised as part of these systems[107-109]. Traditional epidemiological methods struggle to deal with dynamic neighbourhood phenomena such as gentrification, and the relational and adaptive processes by which the residents and resources in neighbourhoods influence each other over time[110]. Complex systems modelling techniques appear to hold considerable promise for examining the multi-dimensional and dynamic interactions that link places – such as neighbourhoods – and the individuals within them, to these outcomes[111].

9.8.2. Potential for using and extending the UK Biobank resource in further research

Considerable further research using the UK Biobank resource could be designed to build on the work of this thesis. With the release of linked primary care data imminent, it will soon be possible to examine other prospective outcomes such as Type 2 diabetes, similar to the way I have used the hospital episodes data in Chapter 8. Following on from Chapter 5's analysis of gene-neighbourhood interactions, the genetic data could be further utilised

to construct risk scores more specific to the neighbourhood exposure (e.g. group together only those SNPs linked specifically to dietary intake). While I have focussed throughout the thesis only on two-way interactions, it may be that important 3-way (or n-way) interactions have been overlooked, bearing in mind challenges for interpretation.

Alongside these suggestions for future research, the process of undertaking this PhD has highlighted for me several ways that the UK Biobank resource could be extended to make it more useful for this kind of research. These include the addition of purpose-designed exposure measures to UK Biobank to reduce the impact of information bias (e.g. exposure misclassification), particularly for measurement of the food environment. Enrichment of the UKBUMP with other data sources could also enable disaggregation of PA facilities to allow greater specificity in exposure definition. While activity space data may not have been feasible on the scale of UK Biobank, one improvement on the street-network buffers and distances used in UKBUMP might have been what Adams et al refer to as "pedestrian-enhanced" network buffers[112], which take into account footpaths in addition to roads. For the distance measures in UKBUMP, reliance on distance to only the single nearest destination of a type (fast-food/takeaway outlet in the case of this thesis) may not be as appropriate as average distance to multiple (e.g. five) nearest destinations of that type[113,114]. Inconsistent results have been reported in comparisons of those two kinds of distance measures of fast-food proximity in relation to BMI[114].

Even more useful would be to add measures of neighbourhood exposure at additional time points, to capture change in exposure over time. Options for doing this include:

a) repeating the existing metric-generation process on updated address points (e.g. 15 years post-baseline);

b) linking small area environmental data to both the baseline address points and updated address points (recognising that small area data rather than ego-centric data are less than ideal, but it may be a less resource-intensive approach than (a));

c) linkage of historical environmental data based on baseline addresses, for the subset of long-term residents at baseline address;

d) additional data requests for historical addresses of cohort members, followed by retrospective ascertainment of pre-baseline exposure.

9.9. Conclusion

In this thesis I have sought to contribute novel evidence about sources of potential effect heterogeneity in relationships between neighbourhood built environments and health, as well as to replicate results of other studies, using a new and very large dataset. The findings highlight effect modification relationships with potentially important policy and public health implications, and which, with further research, will advance our understanding of how, where and for whom neighbourhoods matter for health.

In a very large sample of UK adults at a critical stage of the life course for the development of chronic disease, I investigated associations between three characteristics of the residential neighbourhood environment and obesity- and NCD-related outcomes, with a focus on characterising potential sources of effect heterogeneity operating at multiple scales. In summary, I found that greater availability of formal physical activity facilities near home is associated with lower adiposity across multiple measures, but is less clearly related to lower risk of CVD and cancer outcomes. Importantly, these relationships appear to vary according to several other factors, in particular individual and area-level socioeconomic factors, sex, and other characteristics of the neighbourhood built environment, with population-wide average associations concealing substantial heterogeneity of estimated effects. Living further away from a fast-food outlet was only weakly associated with lower BMI and lower risk of CVD and cancer in the sample as a whole but the association with BMI was substantially modified by sex, and by genetic risk of obesity, with evidence that higher risk individuals may be more sensitive to their food environment. Both exposures were more strongly associated with adiposity outcomes in some parts of England than in others, but further research is needed to identify the factors driving the observed geographical heterogeneity. Neighbourhood greenspace was not associated with any of the examined outcomes across the sample as a whole, but increased greenspace did appear to provide some protection against cancer outcomes in more deprived areas.

The UK Biobank resource provides unique opportunities to explore the complexity of relationships between neighbourhood environments and health in the UK, and in this PhD I have taken some of those opportunities to contribute to knowledge in this field. Extensions to the UK Biobank resource to enrich these opportunities, alongside work in other cohorts, contributions from other disciplines, and targeted intervention research, are needed to further deepen our understanding of these relationships. This thesis highlights the importance of all such work taking into account the study context and the distribution of a range of effect modifiers in the study population.

In their 2015 paper, Keyes & Galea[115(p.308)] argue that "variation in the magnitudes of our associations across time and place are a critical part of the epidemiology of our outcomes, and by drawing on this variation, we may be able to acquire a stronger foothold into how we can shift population health more dramatically in the contexts in which we study". I contend that beyond variation across time and place, we stand to acquire an even deeper understanding – a yet stronger foothold – by investigating variation in the magnitudes of associations across a wider spectrum of factors that might shape the way neighbourhoods influence health. This will contribute to our understanding of how factors operating at multiple levels jointly shape health. I demonstrate in this thesis that the factors that modify neighbourhood effects can span the genetic right through to the macro-environmental forces that influence the way we engage with the world. This advances the field both conceptually and empirically, and has important implications for the translation of research evidence into public health policy, a process that rests on assumptions of generalisability. By better understanding sources of effect heterogeneity, we will be able to better target built environment interventions for health at the places where they will be most effective, and tailor them to benefit the people who stand to, and need to, gain the most.

REFERENCES

1 Egger G, Swinburn B. An 'ecological' approach to the obesity pandemic. *Br Med J* 1997; **315**: 144–80.

2 Dahlgren G, Whitehead M. Policies and strategies to promote social equity in health Background document to WHO Strategy paper for Europe. World Health Organization. 1991.

3 Macintyre S, Ellaway A. Neighborhoods and Health: an overview. In: Kawachi I, Berkman LF, eds. Neighborhoods and Health. Oxford: Oxford University Press, 2003: 20–42.

4 Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev* 2011; **12**: 217–30.

5 Kremers SPJ, de Bruijn G-J, Visscher TLS, van Mechelen W, de Vries NK, Brug J. Environmental influences on energy balance-related behaviors: a dual-process view. *Int J Behav Nutr Phys Act* 2006; **3**: 9.

6 Schneider S, Diehl K, Görig T, *et al.* Contextual influences on physical activity and eating habits: options for action on the community level. *BMC Public Health* 2017; **17**: 760.

7 Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Prev Med* 1999; **29**: 563–70.

8 Ralston J, Brinsden H, Buse K, *et al.* Time for a new obesity narrative. *Lancet* 2018; **392**: 2016–8.

9 Butland B, Jebb S, Kopelman P, *et al.* Foresight Tackling Obesities: Future Choices – project report. London, UK, 2007

10 Rutter H, Savona N, Glonti K, *et al.* The need for a complex systems model of evidence for public health. *Lancet* 2017; **390**: 2602–4.

11 Cummins S, Curtis S, Diez-Roux A V., Macintyre S. Understanding and representing 'place' in health research: a relational approach. *Soc Sci Med* 2007; 65: 1825–38.

12 Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: Off-roading with social epidemiology--exploration, causation, translation. *Am J Epidemiol* 2013; **178**: 858–63.

13 Diez Roux A V. Conceptual approaches to the study of health disparities. *Annu Rev Public Health* 2012; **33**: 41–58.

14 Krieger N. Embodiment: a conceptual glossary for epidemiology. *J Epidemiol Community Health* 2005; **59**: 350–5.

15 Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between *FTO* genotype and *BMI*. *Proc Natl Acad Sci USA* 2015; **112**: 354–9.

16 Robinette JW, Boardman JD, Crimmins EM. Differential vulnerability to neighbourhood disorder: a gene×environment interaction study. *J Epidemiol Community Health* 2019; **73**: 388-92.

17 Mercille G, Richard L, Gauvin L, et al. The food environment and diet quality of urban-dwelling older women and men: assessing the moderating role of diet knowledge. *Can J Public Heal* 2016; **107**: 34.

18 Stafford M, Cummins S, Macintyre S, Ellaway A, Marmot M. Gender differences in the associations between health and neighbourhood environment. *Soc Sci Med* 2005; **60**: 1681–92.

19 Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 2014; **348**: g1464.

20 Sarkar C, Webster C, Gallacher J. Are exposures to ready-to-eat food environments associated with type 2 diabetes? A cross-sectional study of 347 551 UK Biobank adult participants. *Lancet Planet Heal* 2018; **2**: e438–50.

21 Coen SE, Rosenberg MW, Davidson J. "It's gym, like g-y-m not J-i-m": exploring the role of place in the gendering of physical activity. *Soc Sci Med* 2018; **196**: 29–36.

22 CSDH. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. Geneva, 2008.

23 Davey Smith G, Hart C, Watt G, Hole D, Hawthorne V. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study. *J Epidemiol Community Heal* 1998; **52**: 399–405.

24 Adachi-Mejia AM, Lee C, Lee C, *et al.* Geographic variation in the relationship between body mass index and the built environment. *Prev Med* 2017; **100**: 33–40.

25 Clary C, Lewis DJ, Flint E, Smith NR, Kestens Y, Cummins S. The local food environment and fruit and vegetable intake: a geographically weighted regression approach in the ORiEL study. *Am J Epidemiol* 2016; 184: 837–46.

26 Chen M, Creger T, Howard V, Judd SE, Harrington KF, Fontaine KR. Association of community food environment and obesity among US adults: a geographical information system analysis. *J Epidemiol Community Health* 2019; 73: 148–55.

27 Feuillet T, Charreire H, Menai M, *et al.* Spatial heterogeneity of the relationships between environmental characteristics and active commuting: towards a locally varying social ecological model. *Int J Health Geogr* 2015; **14**: 12.

28 Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002; **55**: 125–39.

29 Shelton NJ. Regional risk factors for health inequalities in Scotland and England and the 'Scottish effect'. *Soc Sci Med* 2009; **69**: 761–7.

30 Frohlich KL, Corin E, Potvin L. A theoretical proposal for the relationship between context and disease. *Sociol Heal Illn* 2001; **23**: 776–97.

31 Myers CA, Denstel KD, Broyles ST. The context of context: examining the associations between healthy and unhealthy measures of neighborhood food, physical activity, and social environments. *Prev Med* 2016; **93**: 21–6.

32 Saelens BE, Sallis JF, Frank LD, *et al.* Obesogenic neighborhood environments, child and parent obesity: the neighborhood impact on kids study. *Am J Prev Med* 2012; **42**: e57–64.

33 Boone-Heinonen J, Diez-Roux A V., Goff DC, *et al.* The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study. *PLoS One* 2013; **8:** e85141.

34 Bambra C, Smith KE, Pearce J. Scaling up: The politics of health and place. *Soc Sci Med* 2019; **232**: 36-42.

35 VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Method* 2014; **3**: 33–72.

36 Higgerson J, Halliday E, Ortiz-Nunez A, Brown R, Barr B. Impact of free access to leisure facilities and community outreach on inequalities in physical activity: a quasi-experimental study. *J Epidemiol Community Heal* 2018; **0**: 1–7.

37 Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013; **67**: 190–3.

38 Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. *Int J Behav Nutr Phys Act* 2017; **14**: 156.

39 Shoveller J, Viehbeck S, Di Ruggiero E, Greyson D, Thomson K, Knight R. A critical examination of representations of context within research on population health interventions. *Crit Public Health* 2016; **26**: 487–500.

40 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**: 1026–1034.

41 Rothman KJ, Gallacher JEJ, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol* 2013; **42**: 1012–4.

42 Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. *Int J Epidemiol* 2013; **42**: 1018–22.

43 Richiardi L, Pearce N, Pagano E, Di Cuonzo D, Zugna D, Pizzi C. Baseline selection on a collider: a ubiquitous mechanism occurring in both representative and selected cohort studies. *J Epidemiol Community Health* 2019; **73**: 475-80.

44 Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2017; : 1–10.

45 Ebrahim S, Smith GD. Commentary: Should we always deliberately be non-representative? *Int J Epidemiol* 2013; **42**: 1022–6.

46 Rothman KJ, Hatch EE, Gallacher JEJ. Representativeness is not helpful in studying heterogeneity of effects across subgroups. *Int J Epidemiol* 2014; **43**: 633–4.

47 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010; **340**: 1402–6

48 Smith L, Panter J, Ogilvie D. Characteristics of the environment and physical activity in midlife: findings from UK Biobank. *Prev Med* 2018; **118**: 150–8.

49 Sarkar C. Residential greenness and adiposity: Findings from the UK Biobank. *Environ Int* 2017; **106**: 1–10.

50 Chaix B, Merlo J, Evans D, Leal C, Havard S. Neighbourhoods in eco-epidemiologic research: delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Soc Sci Med* 2009; **69**: 1306–10.

51 Kwan MP. The uncertain geographic context problem. *Ann Assoc Am Geogr* 2012; **102**: 958–68.

52 James P, Hart JE, Arcaya MC, Feskanich D, Laden F, Subramanian S V. Neighborhood self-selection: the role of pre-move health factors on the built and socioeconomic environment. *Int J Environ Res Public Health* 2015; **12**: 12489–504.

53 McCormack GR, Shiell A. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int J Behav Nutr Phys Act* 2011; **8**: 125.

54 Grafova IB, Freedman VA, Lurie N, Kumar R, Rogowski J. The difference-in-difference method: assessing the selection bias in the effects of neighborhood environment on health. *Econ Hum Biol* 2014; **13**: 20–33.

55 Mackenbach JD, Matias de Pinho MG, Faber E, *et al.* Exploring the cross-sectional association between outdoor recreational facilities and leisure-time physical activity: the role of usage and residential self-selection. *Int J Behav Nutr Phys Act* 2018; **15**: 1–11.

56 Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *Int J Behav Nutr Phys Act* 2007; **4**: 32.

57 Lamb KE, Ferguson NS, Wang Y, Ogilvie D, Ellaway A. Distribution of physical activity facilities in Scotland by small area measures of deprivation and urbanicity. *Int J Behav Nutr Phys Act* 2010; **7**: 76.

58 Panter J, Jones A, Hillsdon M. Equity of access to physical activity facilities in an English city. *Prev Med* 2008; **46**: 303–7.

59 Cummins S, Clary C, Shareck M. Enduring challenges in estimating the effect of the food environment on obesity. *Am J Clin Nutr* 2017; **106**: 445–6.

60 Clary CM, Ramos Y, Shareck M, Kestens Y. Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. *Prev Med* 2015; **71**: 83–7.

61 Mason KE, Bentley RJ, Kavanagh AM. Fruit and vegetable purchasing and the relative density of healthy and unhealthy food stores: evidence from an Australian multilevel study. *J Epidemiol Community Health* 2013; **67**: 231–6.

62 Jekanowski MD, Binkley JK, Eales J. Convenience, Accessibility, and the Demand for Fast Food. *J Agric Resour Econ* 2001; **26**: 58–74.

63 Gordon-Larsen P. Food availability/convenience and obesity. *Adv Nutr* 2014; **5**: 809–17.

64 Yee W, Van Arsdol MD. Residential mobility, age, and the life cycle. *Journals Gerontol* 1977; **32**: 211–21.

65 Coulter R, Ham M Van, Coulter R, Ham M Van. Contextualised Mobility Histories of Moving Desires and Actual Moving Behaviour. Institute for the Study of Labor Discussion Papers. 2011; Discussion Paper 6146.

66 Arcaya M, Tucker-Seeley R, Kim R, Schnake-Mahl A, So M, Subramanian SV. Research on neighborhood effects on health in the United States: a systematic review of study characteristics. *Soc Sci Med* 2016; **168**: 16–29.

67 Hernán MA. The C-Word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018; **108**: 616–9.

68 Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Curr Epidemiol Reports* 2016; **3**: 63–71.

69 Oakes JM. Commentary: Advancing neighbourhood-effects research: selection, inferential support, and structural confounding. *Int J Epidemiol* 2006; **35**: 643–7.

70 Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *Am J Epidemiol* 2010; **171**: 664–73.

71 Hernán MA, Vanderweele TJ. Compound treatments and transportability of causal inference. *Epidemiology* 2011; **22**: 268–77.

72 Ou S. Are some neighborhoods bad for your waistline? A test of neighborhood exposure effects on BMI. *J Health Econ* 2018; **63**: 52–63.

73 Jokela M. Are neighborhood health associations causal? A 10-year prospective cohort study with repeated measurements. *Am J Epidemiol* 2014; **180**: 776–84.

74 Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 2004; **58**: 1929–52.

75 Lawlor DA, Tilling K, Smith GD. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; **45**: 1866–86.

76 Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol* 2016; : dyw114.

77 Glymour MM, Rudolph KE. Causal inference challenges in social epidemiology: bias, specificity, and imagination. *Soc Sci Med* 2016; **166**: 258–65.

78 Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* 2016; **45**: 1776–86.

79 Galea S, Link BG. Six paths for the future of social epidemiology. *Am J Epidemiol* 2013; **178**: 843–9.

80 Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Neighbourhood typologies and associations with body mass index and obesity: a cross-sectional study. *Prev Med* 2018; **111**: 351–7.

81 Leal C, Bean K, Thomas F, Chaix B. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol* 2012; **175**: 1152–62.

82 Osypuk TL. Shifting from policy relevance to policy translation: do housing and neighborhoods affect children's mental health? *Soc Psychiatry Psychiatr Epidemiol* 2015; **50**: 215–7.

83 Richardson EA, Mitchell R. Gender differences in relationships between urban green space and health in the United Kingdom. *Soc Sci Med* 2010; **71**: 568–75.

84 Dahlgren G, Whitehead M. European strategies for tackling social inequities in health: Levelling up Part 2. Copenhagen: WHO Regional Office for Europe; 2006.

85 Craig P, Di Ruggiero E, Frohlich KL, Mykhalovskiy E and White M, on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group. Taking account of context in population health intervention research: guidance for producers, users and funders of research. Southampton: NIHR Evaluation, Trials and Studies Coordinating Centre; 2018.

86 Watts P, Phillips G, Petticrew M, Harden A, Renton A. The influence of environmental factors on the generalisability of public health research evidence: physical activity as a worked example. *Int J Behav Nutr Phys Act* 2011; **8**: 128.

87 Bryan AD, Hutchison KE. The role of genomics in health behavior change: challenges and opportunities. *Public Health Genomics* 2012; **15**: 139–45.

88 McBride CM, Bryan AD, Bray MS, Swan GE, Green ED. Health behavior change: can genomics improve behavioral adherence? *Am J Public Health* 2012; **102**: 401–5.

89 Mitchell R, Popham F. Effect of exposure to natural environment on health inequalities: an observational population study. *Lancet* 2008; **372**: 1655–60.

90 Roe J, Aspinall PA, Thompson CW. Understanding relationships between health, ethnicity, place and the role of urban green space in deprived urban communities. *Int J Environ Res Public Health* 2016; **13**.

91 Markevych I, Schoierer J, Hartig T, *et al.* Exploring pathways linking greenspace to health: theoretical and methodological guidance. *Environ Res* 2017; **158**: 301–17.

92 Rigolon A. A complex landscape of inequity in access to urban parks: a literature review. *Landsc Urban Plan* 2016; **153**: 160–9.

93 van Heezik Y, Brymer E. Nature as a commodity: what's good for human health might not be good for ecosystem health. *Front Psychol* 2018; **9**: 1–5.

94 Stuckler D. Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations. *Milbank Q* 2008; **86**: 273-326.

95 Wagner KH, Brath H. A global view on the development of non communicable diseases. *Prev Med* 2012; **54**: S38-41.

96 Pearce JR. Complexity and Uncertainty in Geography of Health Research: Incorporating Life-Course Perspectives. *Ann Am Assoc Geogr* 2018; **4452**: 1–8.

97 Schmidt NM, Nguyen QC, Osypuk TL. Experimental and quasi-experimental designs in neighborhood health effects research: strengthening causal inference and promoting translation. In: Duncan DT, Kawachi I, eds. Neighborhoods and Health, 2nd edn. New York: Oxford University Press, 2018.

98 Blacksher E, Lovasi GS. Place-focused physical activity research, human agency, and social justice in public health: taking agency seriously in studies of the built environment. *Health Place* 2012; **18**: 172–9.

99 Williams GH. The determinants of health: structure, context and agency. *Sociol Health Illn* 2010; **25**: 131–54.

100 Cannuscio CC, Tappe K, Hillier A, Buttenheim A, Karpyn A, Glanz K. Urban food environments and residents' shopping behaviors. *Am J Prev Med* 2013; **45**: 606–14.

101 Cannuscio CC, Hillier A, Karpyn A, Glanz K. The social dynamics of healthy food shopping and store choice in an urban environment. *Soc Sci Med* 2014; **122**: 13–20.

102 Walton E. Vital places: facilitators of behavioral and social health mechanisms in low-income neighborhoods. *Soc Sci Med* 2014; **122**: 1–12.

103 Thompson C, Cummins S, Brown T, Kyle R. Understanding interactions with the food environment: an exploration of supermarket food shopping routines in deprived neighbourhoods. *Health Place* 2013; **19**: 116–23.

104 Keene DE. Qualitative methods and neighborhood health research. In: Duncan DT, Kawac, eds. Neighborhoods and Health, 2nd edn. New York: Oxford University Press, 2018.

105 Kelly MP, Russo F. Causal narratives in public health: the difference between mechanisms of aetiology and mechanisms of prevention in non-communicable diseases. *Sociol Heal Illn* 2018; **40**: 82–99.

106 Naimi AI. Integrating complex systems thinking into epidemiologic research. *Epidemiology* 2016; **27**: 843–7.

107 Huang TT, Drewnosksi A, Kumanyika S, Glass TA. A systems-oriented multilevel framework for addressing obesity in the 21st century. *Prev Chronic Dis* 2009; **6**: A82.

108 Diez Roux A V. Complex systems thinking and current impasses in health disparities research. *Am J Public Health* 2011; **101**: 1627–34.

109 Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *Am J Public Health* 2006; **96**: 452–8.

110 Heaton B, El-Sayed A, Galea S. Agent-based models. In: Duncan DT, Kawachi I, eds. Neighborhoods and Health, 2nd edn. Oxford University Press, 2018.

111 Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. *Annu Rev Public Health* 2012; : 357–76.

112 Adams MA, Frank LD, Schipperijn J, *et al.* International variation in neighborhood walkability, transit, and recreation environments using geographic information systems: the IPEN adult study. *Int J Health Geogr* 2014; **13**: 17.

113 Duncan DT, Kawachi I, Subramanian S V, Aldstadt J, Melly SJ, Williams DR. Examination of how neighborhood definition influences measurements of youths' access to tobacco retailers: a methodological note on spatial misclassification. *Am J Epidemiol* 2013; **179**: 373-81.

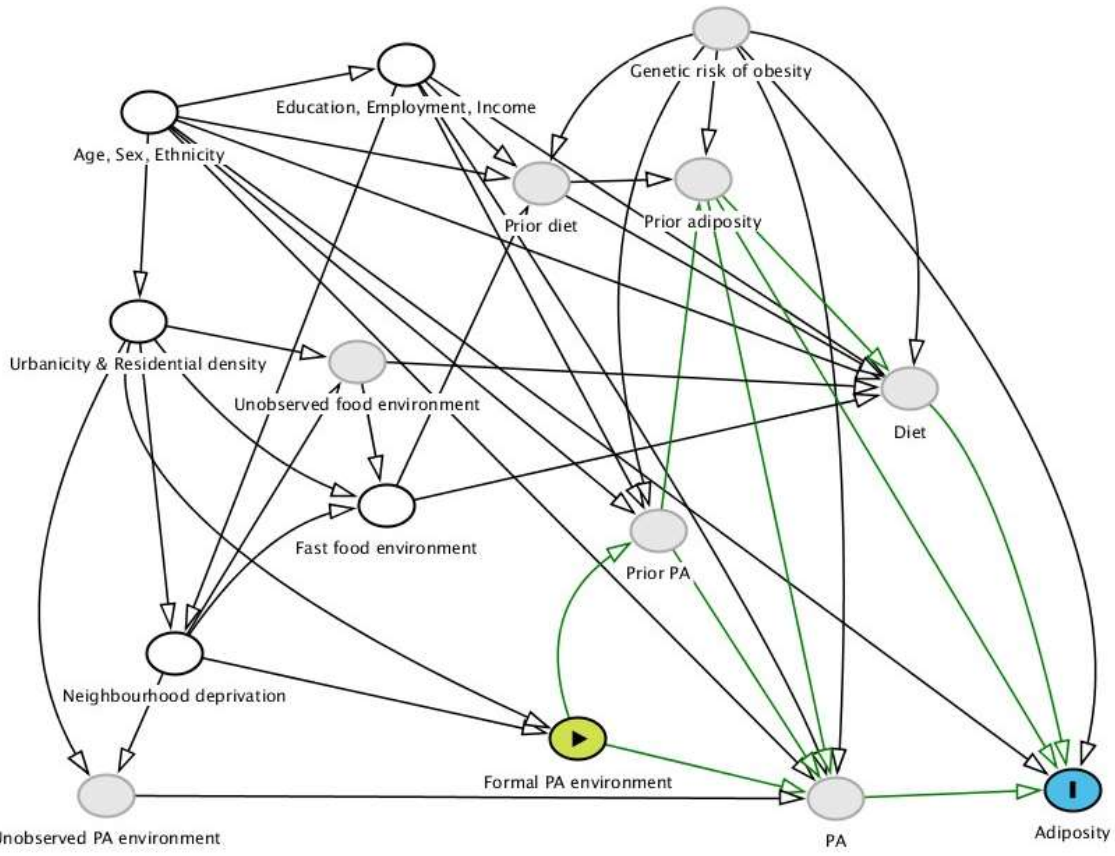
114 Block JP, Christakis NA, O'Malley AJ, Subramanian S V. Proximity to food establishments and body mass index in the framingham heart study offspring cohort over 30 years. *Am J Epidemiol* 2011; **174**: 1108–14.

115 Keyes K, Galea S. What matters most: quantifying an epidemiology of consequence. *Ann Epidemiol* 2015; **25**: 305–11.

Appendix One.

SUPPLEMENTARY MATERIAL: Research Paper 1

1. Directed Acyclic Graphs for Relationships between Neighbourhood Environments (fast-food proximity and density of local formal PA facilities) and Adiposity

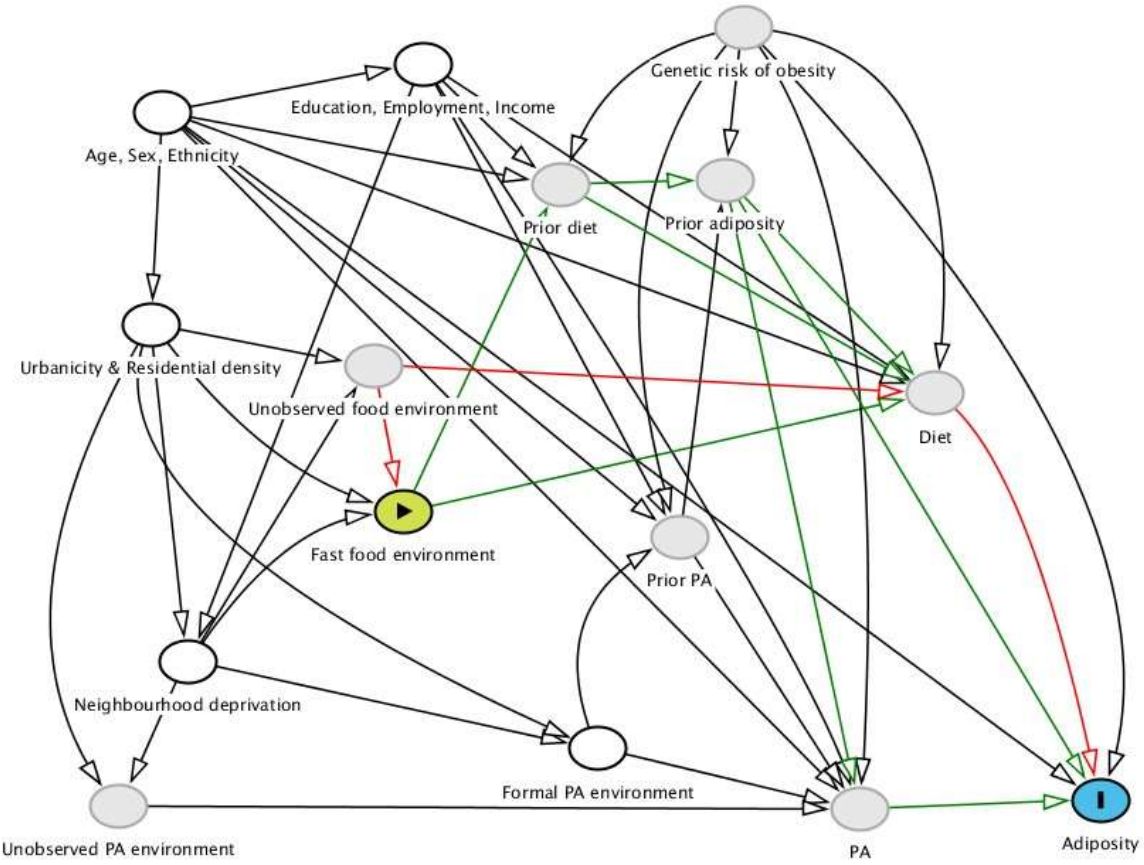


Supplementary Figure 1. Formal physical activity environment – adiposity DAG

Yellow = exposure Blue = outcome White = adjusted potential confounders

Grey = unadjusted or unobserved potential confounders Green line = causal pathway

Red line = potential biasing pathway (absence of red line indicates that adjustment for minimal sufficient set of confounders is achieved)



Supplementary Figure 2. Fast food environment – adiposity DAG

Yellow = exposure Blue = outcome White = adjusted potential confounders
Grey = unadjusted or unobserved potential confounders Green line = causal pathway
Red = potential biasing pathway

2. Classification of Formal Physical Activity Facilities

Formal PA facilities were defined as any land use classified in the Commercial-Leisure subcategory (CL06) of the UK Ordnance Survey AddressBase Premium database (https://www.ordnancesurvey.co.uk/business-and-government/help-and-support/products/addressbase-premium.html). The data are contributed by local authorities, and covers municipal and private facilities for all sporting activities. This subcategory comprises any Indoor/Outdoor Leisure/Sporting Activity/Centre not further defined, as well as the following more specific categories of land use:

- Bowls Facility
- Cricket Facility
- Diving / Swimming Facility
- Equestrian Sports Facility
- Football Facility
- Golf Facility
- Activity / Leisure / Sports Centre
- Playing Field
- Racquet Sports Facility
- Rugby Facility
- Recreation Ground
- Skateboarding Facility
- Civilian Firing Facility
- Tenpin Bowling Facility
- Water Sports Facility
- Winter Sports Facility

Full details of the classification scheme and the types of facilities covered can be found via the link above.

3. Details of Sensitivity Analyses (results summarised in main text)

Adjustment for behavioural confounders (Supplementary Table 1)

Diet is a strong predictor of adiposity but inclusion of total energy intake as a covariate in PA environment-adiposity models would potentially induce selection bias through substantial sample size restriction, as well as confounding bias through other backdoor pathways (Fig 1a, main report) Therefore, we further adjusted PA environment models for dietary intake (using a continuous measure of total energy intake (KJ), based on 24-hour recall dietary assessment). For consistency, additional models of the food environment were adjusted for physical activity. PA was operationalised as self-reported total energy expenditure through physical activity, captured with the self-reported International Physical Activity Questionnaire (IPAQ), expressed in terms of metabolic equivalent (MET) minutes per week, calculated and then categorised (to overcome skewness) according to the IPAQ short form guidelines to reflect low, moderate or high levels of PA.

Sample restriction based on diet and PA data (Supplementary Table 1)

As dietary data were only collected from a subset of 42% of the sample, we also explored whether any effect size attenuation in models adjusted for diet was being driven by selection bias due to missing dietary data, rather than adjustment, by comparing results for the main PA environment models with results from the same model run using only the subsample with dietary data.

PA was missing for 9% of the sample, and for consistency we also ran the fast food environment model on the subsample with PA data, for comparison with the main food environment model.

This enabled us to assess whether any observed differences in estimates when adjusting for behaviours (diet or PA) were being driven by the adjustment (i.e. confounding is present) or by selection bias due to missing data.

Bioimpedance BMI only (Supplementary Table 2)

5580 participants had their BMI calculated from weight measurements taken using standard scales rather than the impedance machine. To test the sensitivity of our results to this, we also estimated models using a version of the BMI measure in which these observations were excluded. Results were almost identical to the primary models.

Sample sizes for the sensitivity analyses varied depending on data completeness for the specific outcome and covariates. Ns are shown in the tables.

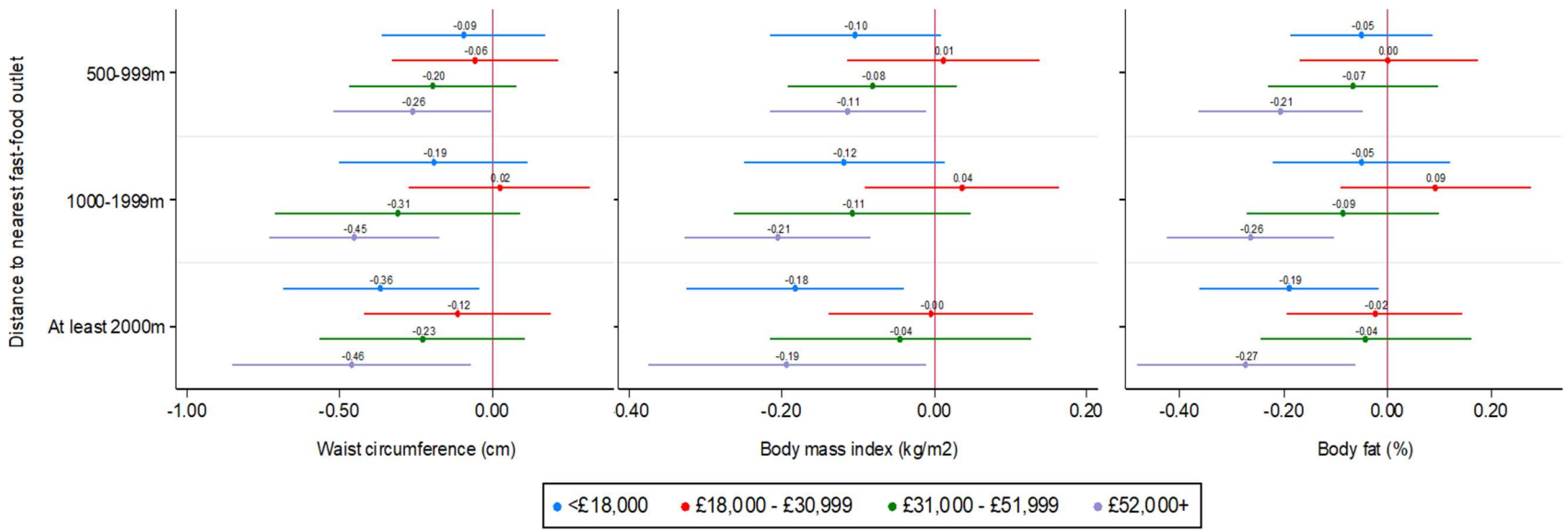
Supplementary Table 1. Results from sensitivity analyses examining impact of adjusting PA environment models for diet and food environment models for physical activity

	Adjusted for diet			*Restricted to subsample with dietary data (but no adjustment for diet)*		
Number of PA resources in 1km street network buffer	**WC**	**BMI (kg/m2)**	**% body fat**	**WC**	**BMI (kg/m2)**	**% body fat**
	n=177,288	n=177,143	n=174,780	n=177,288	n=177,143	n=174,780
0	ref	ref	ref	ref	ref	ref
1	-0.21 (-0.44, 0.01)	-0.11 (-0.20, -0.02)	-0.20 (-0.31, -0.08)	-0.21 (-0.44, 0.01)	-0.11 (-0.20, -0.02)	-0.20 (-0.32, -0.08)
2-3	-0.47 (-0.75, -0.20)	-0.24 (-0.35, -0.13)	-0.32 (-0.48, -0.16)	-0.47 (-0.74, -0.20)	-0.24 (-0.35, -0.13)	-0.32 (-0.49, -0.16)
4-5	-0.82 (-1.18, -0.45)	-0.41 (-0.56, -0.27)	-0.60 (-0.79, -0.41)	-0.81 (-1.18, -0.45)	-0.41 (-0.56, -0.27)	-0.61 (-0.80, -0.41)
6 or more	-1.24 (-1.69, -0.79)	-0.57 (-0.74, -0.41)	-0.83 (-1.05, -0.60)	-1.24 (-1.69, -0.79)	-0.57 (-0.74, -0.41)	-0.83 (-1.05, -0.60)
	Adjusted for PA			*Restricted to subsample with PA data (but no adjustment for PA)*		
Distance to nearest fast-food outlet (m)	**WC**	**BMI (kg/m2)**	**% body fat**	**WC**	**BMI (kg/m2)**	**% body fat**
	n=373,624	n=373,286	n=368,181	n=373,624	n=373,286	n=368,181
<500m	ref	ref	ref	ref	ref	ref
500-999m	-0.18 (-0.33, -0.03)	-0.09 (-0.16, -0.03)	-0.09 (-0.17, -0.01)	-0.17 (-0.32, -0.02)	-0.09 (-0.15, -0.02)	-0.09 (-0.17, 0.00)
1000-1999m	-0.28 (-0.49, -0.06)	-0.12 (-0.22, -0.02)	-0.09 (-0.21, 0.02)	-0.24 (-0.46, -0.02)	-0.11 (-0.21, -0.01)	-0.07 (-0.19, 0.05)
At least 2000m	-0.31 (-0.54, -0.07)	-0.12 (-0.25, 0.02)	-0.13 (-0.27, 0.02)	-0.27 (-0.52, -0.01)	-0.10 (-0.24, 0.04)	-0.10 (-0.25, 0.05)

Supplementary Table 2. Results from sensitivity analyses excluding BMI measurements taken using standard scales rather than bioimpedance machine

	Impedance-only BMI
Number of PA resources in 1km street network buffer	**BMI (kg/m2)**
	n=395,855
0	ref
1	-0.08 (-0.15, 0.00)
2-3	-0.18 (-0.28, -0.07)
4-5	-0.33 (-0.47, -0.20)
6 or more	-0.56 (-0.73, -0.39)
Distance to nearest fast-food outlet (m)	n=395,855
<500m	ref
500-999m	-0.08 (-0.15, -0.02)
1000-1999m	-0.11 (-0.20, -0.01)
At least 2000m	-0.10 (-0.24, 0.04)

4. Fast-food environment and adiposity: effect modification by annual household income



Supplementary Figure 3. Association between distance to nearest fast-food outlet and adiposity, by annual household income

Figure shows annual-household-income-stratified, fully adjusted mean differences in adiposity and associated 95% confidence intervals. The red line at zero represents the reference category (living <500m from nearest fast-food outlet).

5. Standardised coefficients for direct comparison of adiposity and negative control outcomes

The negative control analysis for the PA environment (but not the fast-food environment) in the published paper above indicates possible residual confounding, because an unexpected association with height was observed. As stated in the paper, but not shown, comparison of standardised coefficients across the primary models and the negative control model suggest this residual confounding would only partially account for the observed association between the availability of PA facilities and adiposity. These results are shown in Supplementary Table 3) below. The magnitude of the standardised point estimates for height are not as large as for the adiposity outcomes, especially BMI, for which the standardised estimates are twice the magnitude of those for height.

Supplementary Table 3. Standardised coefficients

	WC (cm)	BMI (kg/m2)	% body fat	Height (%) (negative control)
Number of PA resources in 1km street network buffer	(n=401,917)	(n=401,435)	(n=395,640)	(n=401,675)
0	ref	ref	ref	ref
1	-0.01 (-0.02, 0.00)	-0.02 (-0.03, 0.00)	-0.01 (-0.02, 0.00)	0.01 (0.00, 0.02)
2-3	-0.03 (-0.05, -0.01)	-0.04 (-0.06, -0.02)	-0.03 (-0.05, -0.01)	0.02 (0.01, 0.03)
4-5	-0.05 (-0.07, -0.02)	-0.07 (-0.10, -0.04)	-0.06 (-0.08, -0.03)	0.04 (0.03, 0.05)
6 or more	-0.09 (-0.12, -0.06)	-0.12 (-0.16, -0.08)	-0.10 (-0.12, -0.07)	0.06 (0.04, 0.07)

Appendix Two.

SUPPLEMENTARY MATERIAL: Research Paper 2

Supplementary Table 4. SNPs included in each polygenic risk score

SNP	Chr	Position	Gene	BMI-increasing allele	Effect size (β coefficient per effect allele, SD units of BMI)	Included in 69-SNP risk score	Included in 91-SNP risk score	Reason for exclusion[1,2]
rs1558902	16	52,361,075	*FTO*	A	0.0818	Yes	Yes	
rs17024393	1	109,956,211	*GNAT2*	C	0.0658	Yes	Yes	
rs13021737	2	622,348	*TMEM18*	G	0.0601	Yes	Yes	
rs6567160	18	55,980,115	*MC4R*	C	0.0556	Yes	Yes	
rs11847697	14	29,584,863	*PRKD1*	T	0.0492	Yes	Yes	
rs16851483	3	142,758,126	*RASA2*	T	0.0483	Yes	Yes	
rs543874	1	176,156,103	*SEC16B*	G	0.0482	Yes	Yes	
rs13107325	4	103,407,732	*SLC39A8*	T	0.0477	No	No	Possible pleiotropy
rs1516725	3	187,306,698	*ETV5*	C	0.0451	Yes	Yes	
rs2207139	6	50,953,449	*TFAP2B*	G	0.0447	Yes	Yes	
rs11030104	11	27,641,093	*BDNF*	A	0.0414	No	No	Possible pleiotropy
rs12446632	16	19,842,890	*GPRC5B*	G	0.0403	Yes	Yes	
rs10938397	4	44,877,284	*GNPDA2*	G	0.0402	Yes	Yes	
rs7899106	10	87,400,884	*GRID1*	G	0.0395	Yes	Yes	
rs2287019	19	50,894,012	*QPCTL*	C	0.0360	Yes	Yes	
rs11727676	4	145,878,514	*HHIP*	T	0.0358	Yes	Yes	
rs16907751	8	81,538,012	*ZBTB10*	C	0.0350	No	Yes	Identified in secondary meta-analyses only
rs12429545	13	53,000,207	*OLFM4*	A	0.0334	Yes	Yes	
rs3101336	1	72,523,773	*NEGR1*	C	0.0334	Yes	Yes	
rs2245368	7	76,446,079	*DTX2P1*	C	0.0317	Yes	Yes	
rs7138803	12	48,533,735	*BCDIN3D*	A	0.0315	Yes	Yes	

SNP	Chr	Position	Gene	BMI-increasing allele	Effect size (β coefficient per effect allele, SD units of BMI)	Included in 69-SNP risk score	Included in 91-SNP risk score	Reason for exclusion[1,2]
rs16951275	15	65,864,222	*MAP2K5*	T	0.0311	Yes	Yes	
rs3888190	16	28,796,987	*ATP2A1*	A	0.0309	No	No	Possible pleiotropy
rs11191560	10	104,859,028	*NT5C2*	C	0.0308	Yes	Yes	
rs10182181	2	25,003,800	*ADCY3*	G	0.0307	Yes	Yes	
rs11057405	12	121,347,850	*CLIP1*	G	0.0307	Yes	Yes	
rs17001654	4	77,348,592	*SCARB2*	G	0.0306	No	No	Linkage disequilibrium
rs9581854	13	26,915,782	*MTIF3*	T	0.0298	Yes	Yes	
rs13078960	3	85,890,280	*CADM2*	G	0.0297	Yes	Yes	
rs3810291	19	52,260,843	*ZC3H4*	A	0.0283	Yes	Yes	
rs13191362	6	162,953,340	*PARK2*	A	0.0277	Yes	Yes	
rs3817334	11	47,607,569	*MTCH2*	T	0.0262	Yes	Yes	
rs2112347	5	75,050,998	*POC5*	T	0.0261	Yes	Yes	
rs2075650	19	50,087,459	*TOMM40*	A	0.0258	No	No	Linkage disequilibrium
rs10968576	9	28,404,339	*LINGO2*	G	0.0249	Yes	Yes	
rs17094222	10	102,385,430	*HIF1AN*	C	0.0249	Yes	Yes	
rs2121279	2	142,759,755	*LRP1B*	T	0.0245	Yes	Yes	
rs12566985	1	74,774,781	*FPGT*	G	0.0242	Yes	Yes	
rs7141420	14	78,969,207	*NRXN3*	T	0.0235	Yes	Yes	
rs7903146	10	114,748,339	*TCF7L2*	C	0.0234	No	Yes	Not included in Tyrrell et al's GRS (reason unclear)
rs13201877	6	137,717,234	*IFNGR1*	G	0.0233	No	Yes	Identified in secondary meta-analyses only
rs10132280	14	24,998,019	*STXBP6*	C	0.0230	Yes	Yes	
rs1016287	2	59,159,129	*LINC01122*	T	0.0229	Yes	Yes	
rs657452	1	49,362,434	*AGBL4*	A	0.0227	Yes	Yes	
rs758747	16	3,567,359	*NLRC3*	T	0.0225	Yes	Yes	
rs17405819	8	76,969,139	*HNF4G*	T	0.0224	Yes	Yes	
rs205262	6	34,671,142	*C6orf106*	G	0.0221	Yes	Yes	

SNP	Chr	Position	Gene	BMI-increasing allele	Effect size (β coefficient per effect allele, SD units of BMI)	Included in 69-SNP risk score	Included in 91-SNP risk score	Reason for exclusion[1,2]
rs7599312	2	213,121,476	*ERBB4*	G	0.0220	Yes	Yes	
rs11165643	1	96,696,685	*PTBP2*	T	0.0218	Yes	Yes	
rs12286929	11	114,527,614	*CADM1*	G	0.0217	Yes	Yes	
rs7243357	18	55,034,299	*GRP*	T	0.0217	Yes	Yes	
rs12401738	1	78,219,349	*FUBP1*	A	0.0211	Yes	Yes	
rs17203016	2	207,963,763	*CREB1*	G	0.0210	No	Yes	Identified in secondary meta-analyses only
rs4256980	11	8,630,515	*TRIM66*	G	0.0209	Yes	Yes	
rs11126666	2	26,782,315	*KCNK3*	A	0.0207	Yes	Yes	
rs12885454	14	28,806,589	*PRKD1*	C	0.0207	Yes	Yes	
rs2650492	16	28,240,912	*SBK1*	A	0.0207	Yes	Yes	
rs1167827	7	75,001,105	*HIP1*	G	0.0202	Yes	Yes	
rs9914578	17	1,951,886	*SMG6*	G	0.0201	No	Yes	Identified in secondary meta-analyses only
rs2365389	3	61,211,502	*FHIT*	C	0.0200	Yes	Yes	
rs2176598	11	43,820,854	*HSD17B12*	T	0.0198	Yes	Yes	
rs1460676	2	164,275,935	*FIGN*	C	0.0197	No	Yes	Identified in secondary meta-analyses only
rs2820292	1	200,050,910	*NAV1*	C	0.0195	Yes	Yes	
rs17724992	19	18,315,825	*PGPEP1*	A	0.0194	Yes	Yes	
rs1000940	17	5,223,976	*RABEP1*	G	0.0192	Yes	Yes	
rs2033732	8	85,242,264	*RALYL*	C	0.0192	Yes	Yes	
rs9925964	16	31,037,396	*KAT8*	A	0.0192	No	No	Linkage disequilibrium
rs9641123	7	93,035,668	*CALCR*	C	0.0191	No	Yes	Identified in secondary meta-analyses only
rs2033529	6	40,456,631	*TDRG1*	G	0.0190	No	Yes	Not included in Tyrrell et al's GRS because unavailable
rs1928295	9	119,418,304	*TLR4*	T	0.0188	Yes	Yes	
rs3849570	3	81,874,802	*GBE1*	A	0.0188	Yes	Yes	
rs6091540	20	50,521,269	*ZFP64*	C	0.0188	No	Yes	Identified in secondary meta-analyses only

SNP	Chr	Position	Gene	BMI-increasing allele	Effect size (β coefficient per effect allele, SD units of BMI)	Included in 69-SNP risk score	Included in 91-SNP risk score	Reason for exclusion[1,2]
rs9400239	6	109,084,356	*FOXO3*	C	0.0188	Yes	Yes	
rs9374842	6	120,227,364	*LOC285762*	T	0.0187	No	Yes	Identified in secondary meta-analyses only
rs6804842	3	25,081,441	*RARB*	G	0.0185	Yes	Yes	
rs12940622	17	76,230,166	*RPTOR*	G	0.0182	Yes	Yes	
rs29941	19	39,001,372	*KCTD15*	G	0.0182	Yes	Yes	
rs7164727	15	70,881,044	*LOC100287559*	T	0.0180	No	Yes	Identified in secondary meta-analyses only
rs4740619	9	15,624,326	*C9orf93*	T	0.0179	Yes	Yes	
rs1528435	2	181,259,207	*UBE2E3*	T	0.0178	Yes	Yes	
rs11583200	1	50,332,407	*ELAVL4*	C	0.0177	Yes	Yes	
rs3736485	15	49,535,902	*DMXL2*	A	0.0176	Yes	Yes	
rs1441264	13	78,478,920	*MIR548A2*	A	0.0175	No	Yes	Identified in secondary meta-analyses only
rs10733682	9	128,500,735	*LMX1B*	A	0.0174	Yes	Yes	
rs6477694	9	110,972,163	*EPB41L4B*	C	0.0174	Yes	Yes	
rs11688816	2	62,906,552	*EHBP1*	G	0.0172	Yes	Yes	
rs9540493	13	65,103,705	*MIR548X2*	A	0.0172	No	Yes	Identified in secondary meta-analyses only
rs2080454	16	47,620,091	*CBLN1*	C	0.0168	No	Yes	Identified in secondary meta-analyses only
rs1808579	18	19,358,886	*C18orf8*	C	0.0167	Yes	Yes	
rs977747	1	47,457,264	*TAL1*	T	0.0167	No	Yes	Identified in secondary meta-analyses only
rs6465468	7	95,007,450	*ASB4*	T	0.0166	No	Yes	Identified in secondary meta-analyses only
rs2836754	21	39,213,610	*ETS2*	C	0.0164	No	Yes	Identified in secondary meta-analyses only
rs7239883	18	38,401,669	*LOC284260*	G	0.0164	No	Yes	Identified in secondary meta-analyses only
rs7715256	5	153,518,086	*GALNT10*	G	0.0163	No	Yes	Identified in secondary meta-analyses only
rs4787491	16	29,922,838	*INO80E*	G	0.0159	No	Yes	Identified in secondary meta-analyses only
rs492400	2	219,057,996	*USP37*	C	0.0158	No	Yes	Identified in secondary meta-analyses only
rs2176040	2	226,801,046	*LOC646736*	A	0.0141	No	Yes	Identified in secondary meta-analyses only

Note: This table is derived from Locke et al (2015), with additional information from Tyrrell et al (2017).

Supplementary Table 5. Results of sensitivity analyses using unweighted genetic risk scores

	69-SNP unweighted imputed GRS			*91-SNP unweighted imputed GRS*		
	Quintile of genetic risk	Mean BMI difference for unit increase in neighbourhood exposure	P-interaction	Quintile of genetic risk	Mean BMI difference for unit increase in neighbourhood exposure	P-interaction
Fast-food proximity	Q1	-0.090 (-0.223, 0.043)		Q1	-0.102 (-0.235, 0.031)	
	Q2	-0.119 (-0.243, 0.005)		Q2	-0.124 (-0.248, 0.000)	
	Q3	-0.137 (-0.259, -0.014)	0.070	Q3	-0.137 (-0.260, -0.015)	0.171
	Q4	-0.155 (-0.278, -0.031)		Q4	-0.151 (-0.274, -0.027)	
	Q5	-0.184 (-0.316, -0.051)		Q5	-0.173 (-0.305, -0.041)	
PA facilities	Q1	-0.074 (-0.101, -0.048)		Q1	-0.076 (-0.103, -0.050)	
	Q2	-0.075 (-0.101, -0.050)		Q2	-0.076 (-0.102, -0.051)	
	Q3	-0.076 (-0.102, -0.050)	0.700	Q3	-0.076 (-0.102, -0.051)	0.981
	Q4	-0.076 (-0.102, -0.051)		Q4	-0.076 (-0.102, -0.051)	
	Q5	-0.077 (-0.104, -0.051)		Q5	-0.076 (-0.103, -0.050)	

Supplementary Table 6. Results of sensitivity analyses using an expanded sample including observations from UK Biobank participants of non-White ethnicities

		91-SNP weighted imputed GRS (N=393,993)	
	Quintile of genetic risk	Mean BMI difference for unit increase in neighbourhood exposure	P-interaction
Fast-food proximity	Q1	-0.086 (-0.208, 0.036)	
(log (base 10) of distance (m) to nearest fast-food outlet)	Q2	-0.119 (-0.233, -0.005)	
	Q3	-0.140 (-0.252, -0.028)	0.042
	Q4	-0.161 (-0.274, -0.047)	
	Q5	-0.195 (-0.318, -0.073)	
Availability of PA facilities	Q1	-0.068 (-0.095, -0.041)	
(number of formal PA facilities within 1km of home address)	Q2	-0.071 (-0.097, -0.046)	
	Q3	-0.072 (-0.098, -0.046)	0.270
	Q4	-0.074 (-0.100, -0.048)	
	Q5	-0.077 (-0.104, -0.050)	

Supplementary Table 7. Results of sensitivity analyses additionally adjusting for extra 30 genetic ancestry principal components and birth location

	Fast-food proximity		**Availability of PA facilities**	
P-for-interaction	Adjusted for 40 ancestral PCs	Adjusted for 40 PCs & place of birth	Adjusted for 40 ancestral PCs	Adjusted for 40 PCs & place of birth
69-SNP GRS	0.015	0.030	0.385	0.395
91-SNP GRS	0.028	0.069	0.536	0.550
rs1558902 (FTO)	0.066	0.088	0.928	0.904
rs6567160 (MC4R)	0.009	0.025	0.617	0.766
rs13021737 (TMEM18)	0.999	0.967	0.081	0.063
rs13078960 (CADM2)	0.108	0.142	0.407	0.361
rs10938397 (GNPDA2)	0.306	0.301	0.732	0.964
rs7141420 (NRXN3)	0.519	0.665	0.114	0.155

REFERENCES

1. Tyrrell J, Wood AR, Ames RM, *et al.* Gene-obesogenic environment interactions in the UK Biobank study. *Int J Epidemiol* 2017; **46**: 559–75

2. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.

Appendix Three.

SUPPLEMENTARY MATERIAL: Research Paper 3

1. Classification of Formal Physical Activity Facilities

Formal PA facilities were defined as any land use classified in the Commercial-Leisure subcategory (CL06) of the UK Ordnance Survey AddressBase Premium database (https://www.ordnancesurvey.co.uk/business-and-government/help-and-support/products/addressbase-premium.html). The data are contributed by local authorities, and covers municipal and private facilities for all sporting activities. This subcategory comprises any Indoor/Outdoor Leisure/Sporting Activity/Centre not further defined, as well as the following more specific categories of land use:

- Bowls Facility
- Cricket Facility
- Diving / Swimming Facility
- Equestrian Sports Facility
- Football Facility
- Golf Facility
- Activity / Leisure / Sports Centre
- Playing Field
- Racquet Sports Facility
- Rugby Facility
- Recreation Ground
- Skateboarding Facility
- Civilian Firing Facility
- Tenpin Bowling Facility
- Water Sports Facility
- Winter Sports Facility

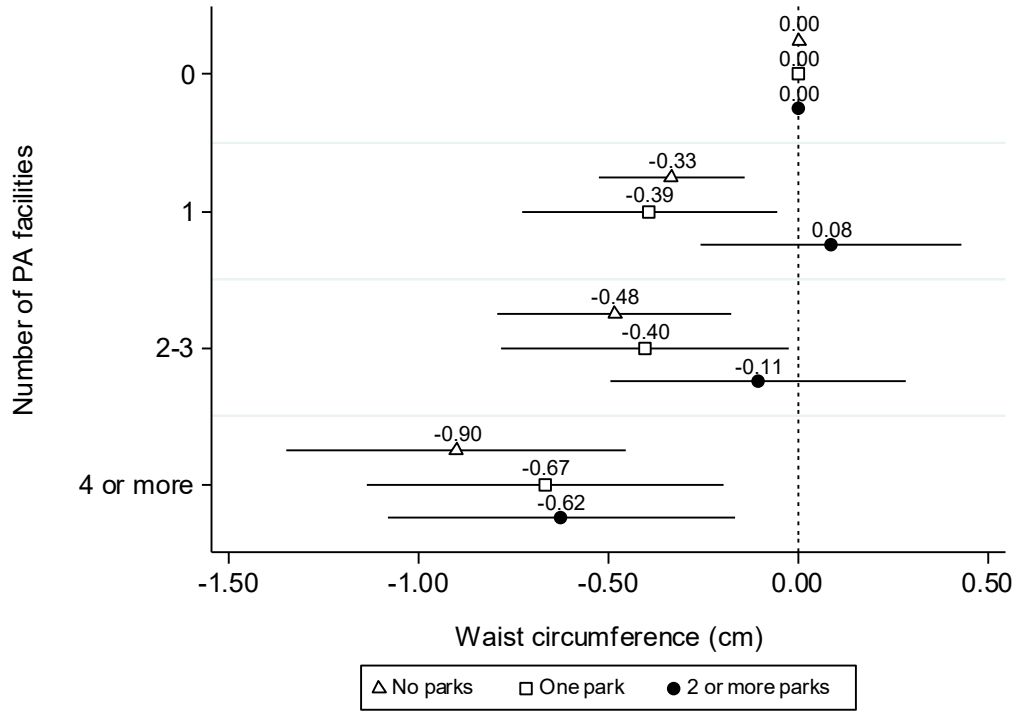
Full details of the classification scheme and the types of facilities covered can be found via the link above.

2. Sensitivity analysis: Alternative adiposity measures

To check for consistency across alternative measures of adiposity, we repeated the primary analyses for waist circumference and body fat percentage outcome measures rather than BMI.

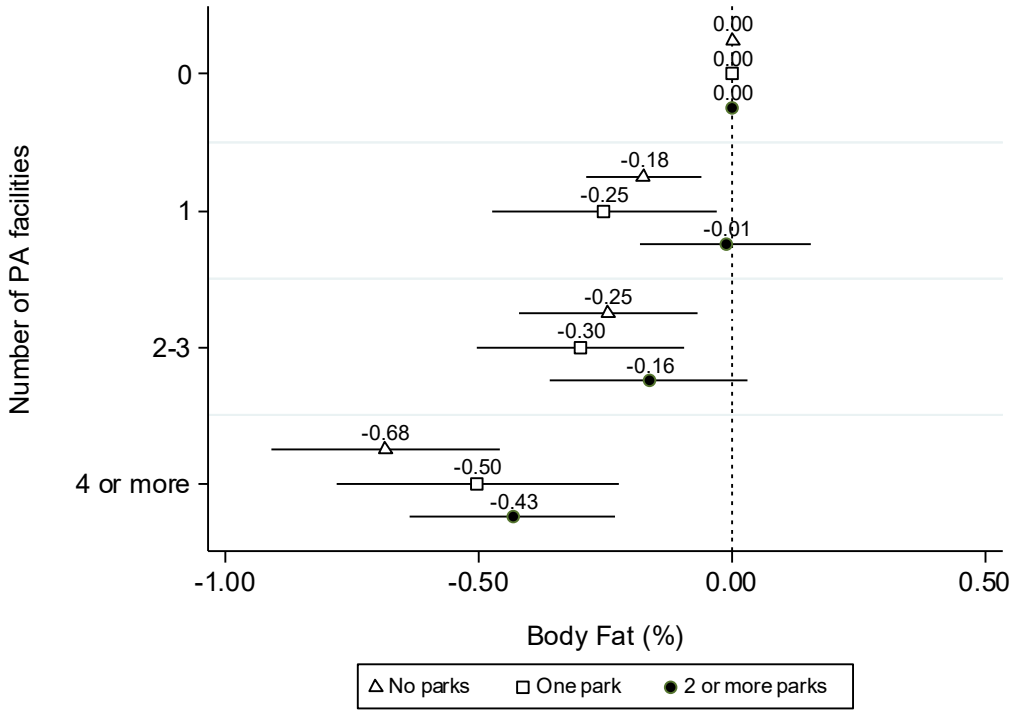
Effect modification by urban park availability

As for BMI, there is some evidence that the number of parks within one kilometre of a person's home modifies the association between the formal PA environment and both waist circumference ($P_{interaction}$ = 0.073) and body far percentage ($P_{interaction}$ = 0.095). Supplementary Figures 4 and 5 show estimates of these associations within strata of urban park availability. Stratification shows that the associations with the alternative measures of adiposity, just as for BMI, is weak in the areas with the most parks, whereas there is a clear inverse association between density of formal PA facilities and both waist circumference and body fat percentage in the areas with no parks.



Supplementary Figure 4. Association between number of formal PA facilities and waist circumference, stratified by number of parks and other public open/green spaces

Figure shows park availability-stratified, fully adjusted mean differences in waist circumference and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and fast-food store proximity.

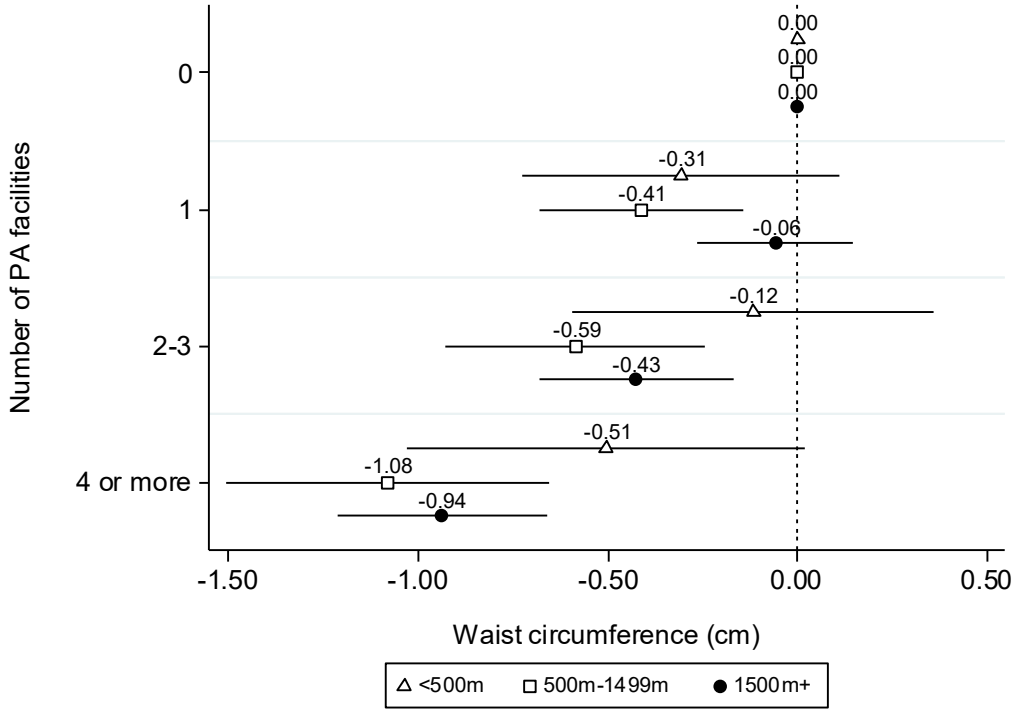


Supplementary Figure 5. Association between number of formal PA facilities and body fat percentage, stratified by number of parks and other public open/green spaces

Figure shows park availability-stratified, fully adjusted mean differences in body fat percentage and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and fast-food store proximity.

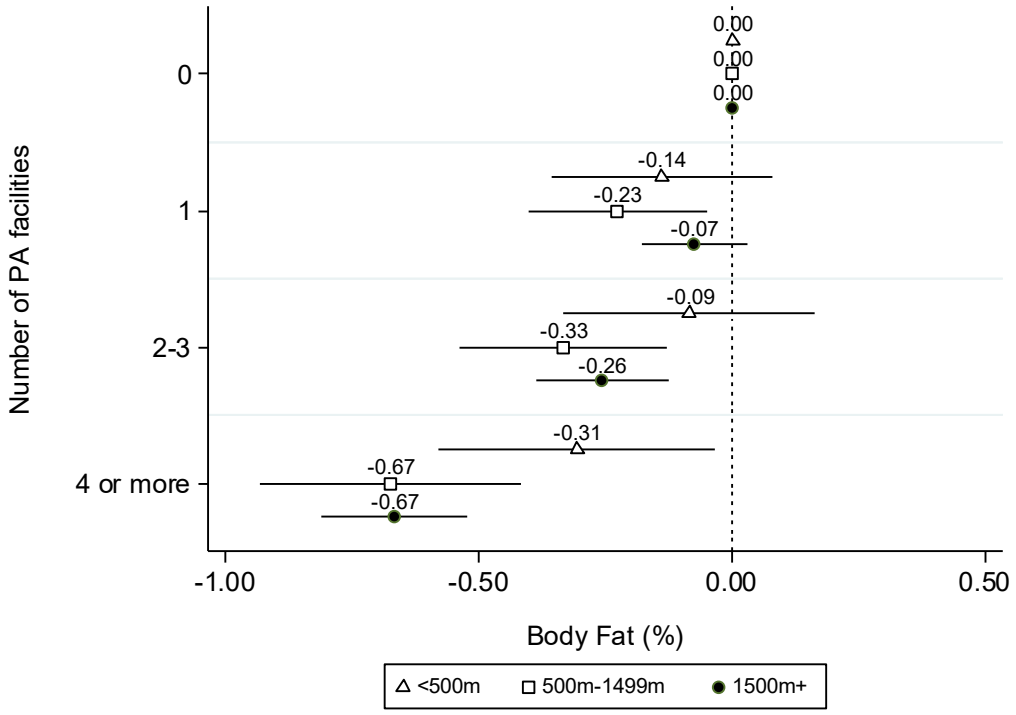
Effect modification by fast-food proximity

As we observed for BMI, there is also strong statistical evidence that associations between the PA environment and both alternative adiposity measures are modified by proximity to a fast-food store ($P_{interaction}$ < 0.0001 for both waist circumference and body fat percentage). Stratified results showed that among people living within 500m of a fast-food store, the association between number of nearby formal PA facilities and these other adiposity measures is considerably less apparent than it is among those who live further from a fast-food store (Supplementary Figures 6 and 7).



Supplementary Figure 6. Association between number of formal PA facilities and waist circumference, stratified by distance to nearest fast-food store

Figure shows fast-food proximity-stratified, fully adjusted mean differences in waist circumference and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and urban park availability.

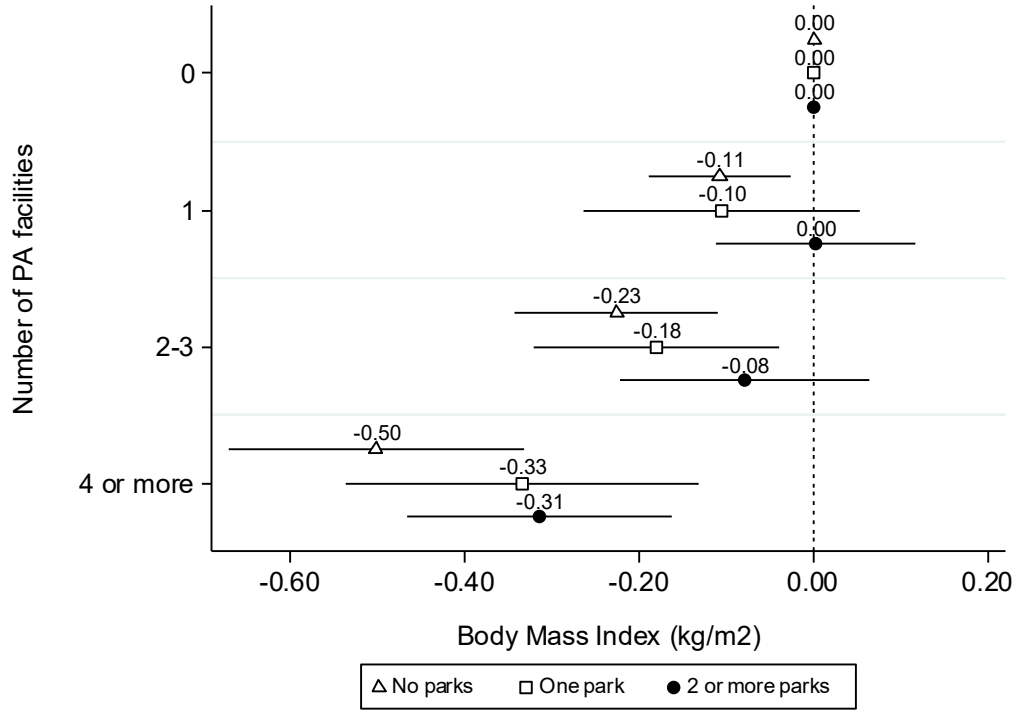


Supplementary Figure 7. Association between number of formal PA facilities and body fat percentage, stratified by distance to nearest fast-food store

Figure shows fast-food proximity-stratified, fully adjusted mean differences in body fat percentage and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, and urban park availability.

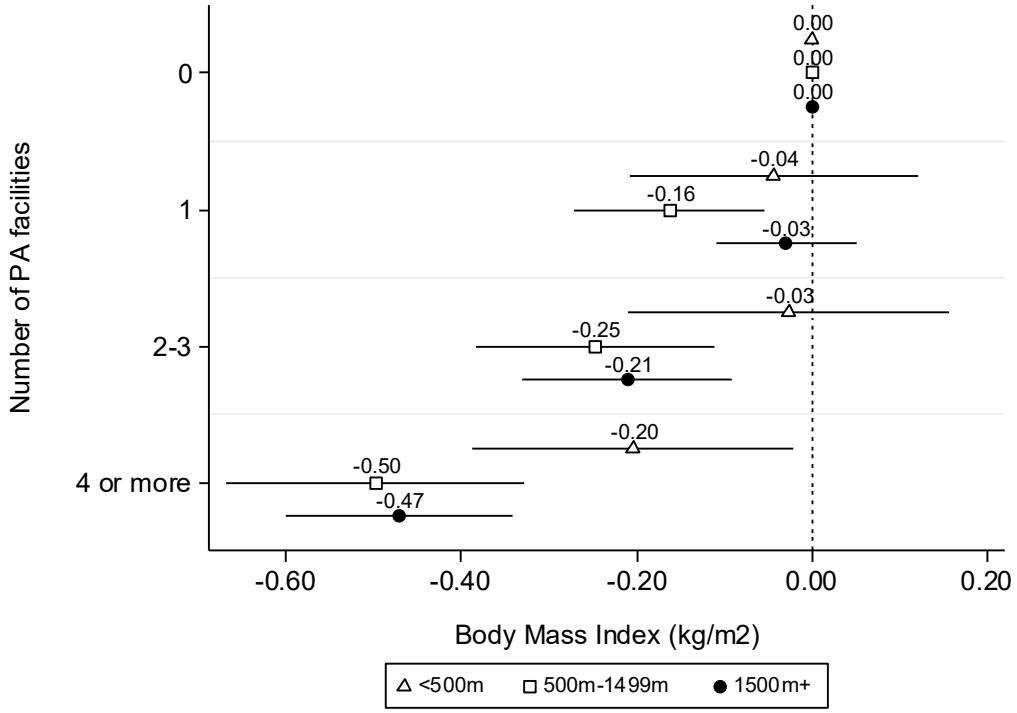
3. Sensitivity analysis: Including non-urban residents and adjusting for urban/non-urban status

To examine the impact of our decision to exclude non-urban participants rather than adjust models for urbanicity, we repeated the primary analysis on the full urban and non-urban sample combined, adjusting for urban/non-urban status. The results of this sensitivity analysis showed that the patterns across both sets of modifier-stratified models (Supplementary Figures 8 and 9) mirrored those observed in the urban-only sample. Statistical evidence of an interaction with fast-food proximity remained strong ($P_{interaction}$ <0.0001) but was weaker for the interaction with park availability ($P_{interaction}$ = 0.235) than it had been in the urban-only sample.



Supplementary Figure 8. Association between number of formal PA facilities and BMI, stratified by park availability, including non-urban residents and adjusting for urban/non-urban status

Figure shows park availability-stratified, fully adjusted mean differences in BMI and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, fast-food store proximity and urban/non-urban status.

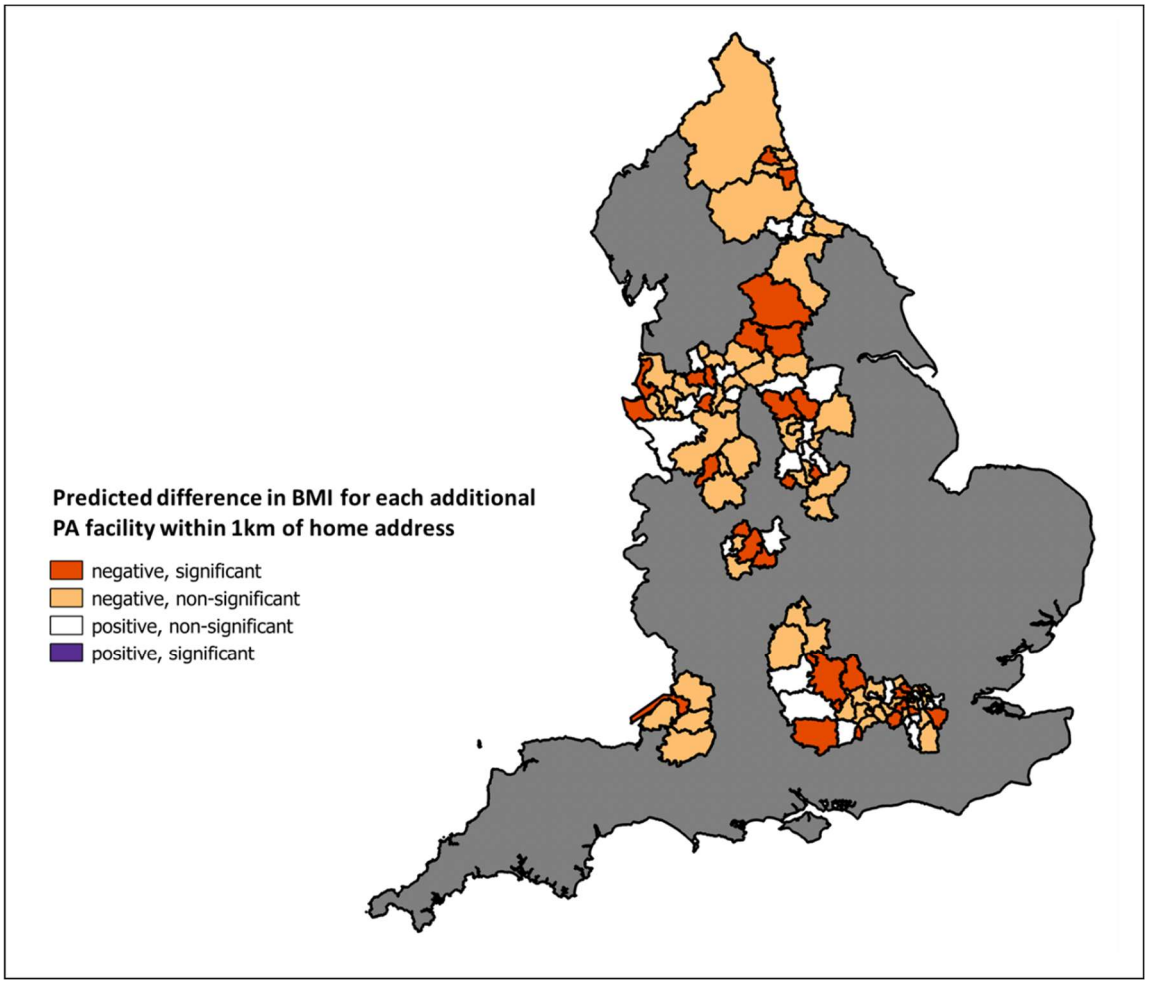


Supplementary Figure 9. Association between number of formal PA facilities and BMI, stratified by distance to nearest fast-food store, including non-urban residents and adjusting for urban/non-urban status

Figure shows fast-food proximity-stratified, fully adjusted mean differences in BMI and associated 95% CIs from multilevel linear regression models. The dashed line at zero represents the reference category (no physical activity facilities with 1km of home). Models are adjusted for age, sex, ethnicity, area deprivation, individual socioeconomic characteristics (income, education, and employment status), residential density, park availability and urban/non-urban status.

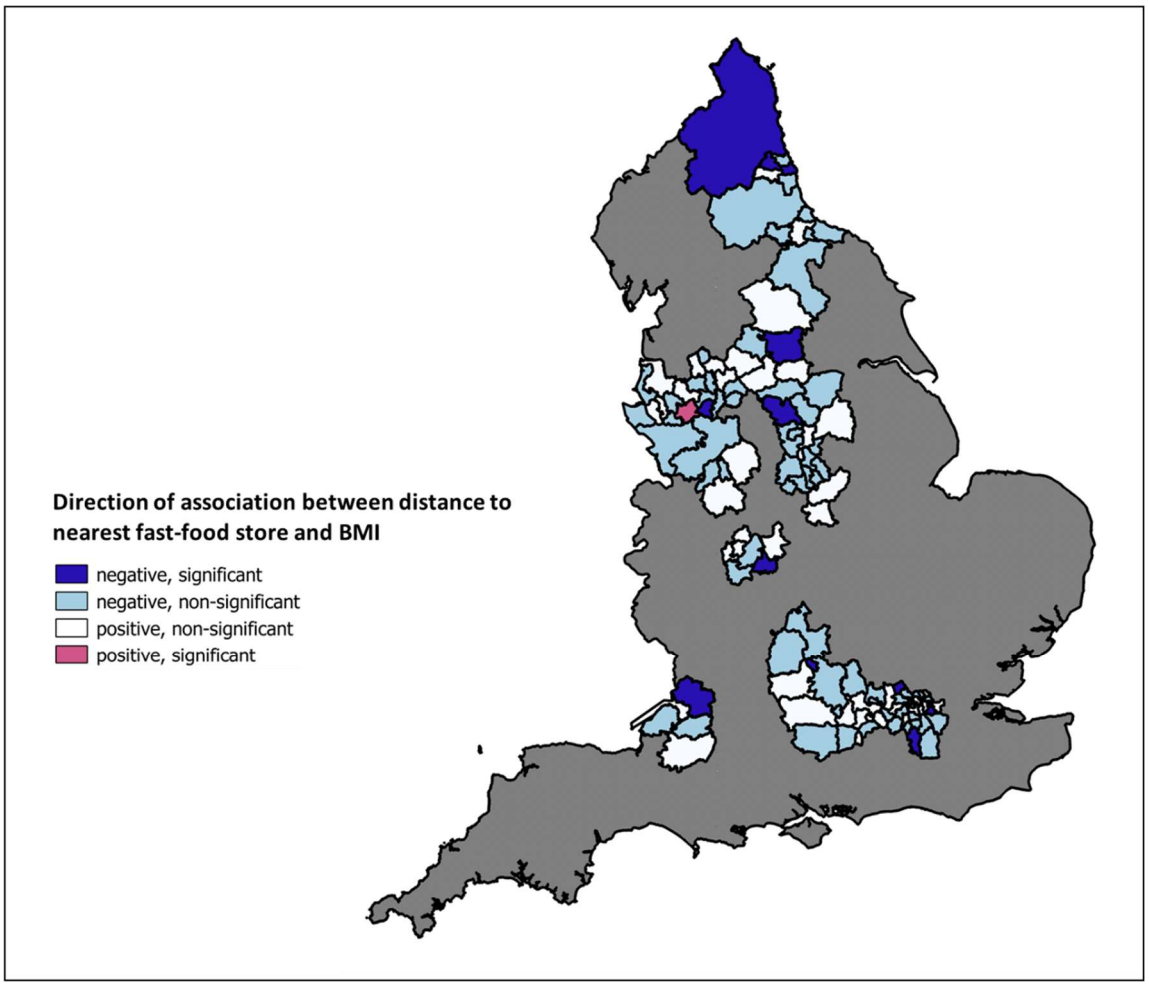
Appendix Four.

SUPPLEMENTARY MATERIAL: Research Paper 4



Supplementary Figure 10. Geographical heterogeneity in the association between availability of formal PA facilities and BMI, from separate regression models of 122 Local Authority Districts of England represented in UK Biobank

Notes: Significance refers to arbitrary threshold of P<0.05. Grey areas in the map were not included in UK Biobank, or were non-urban, or had <200 study participants in the LAD.



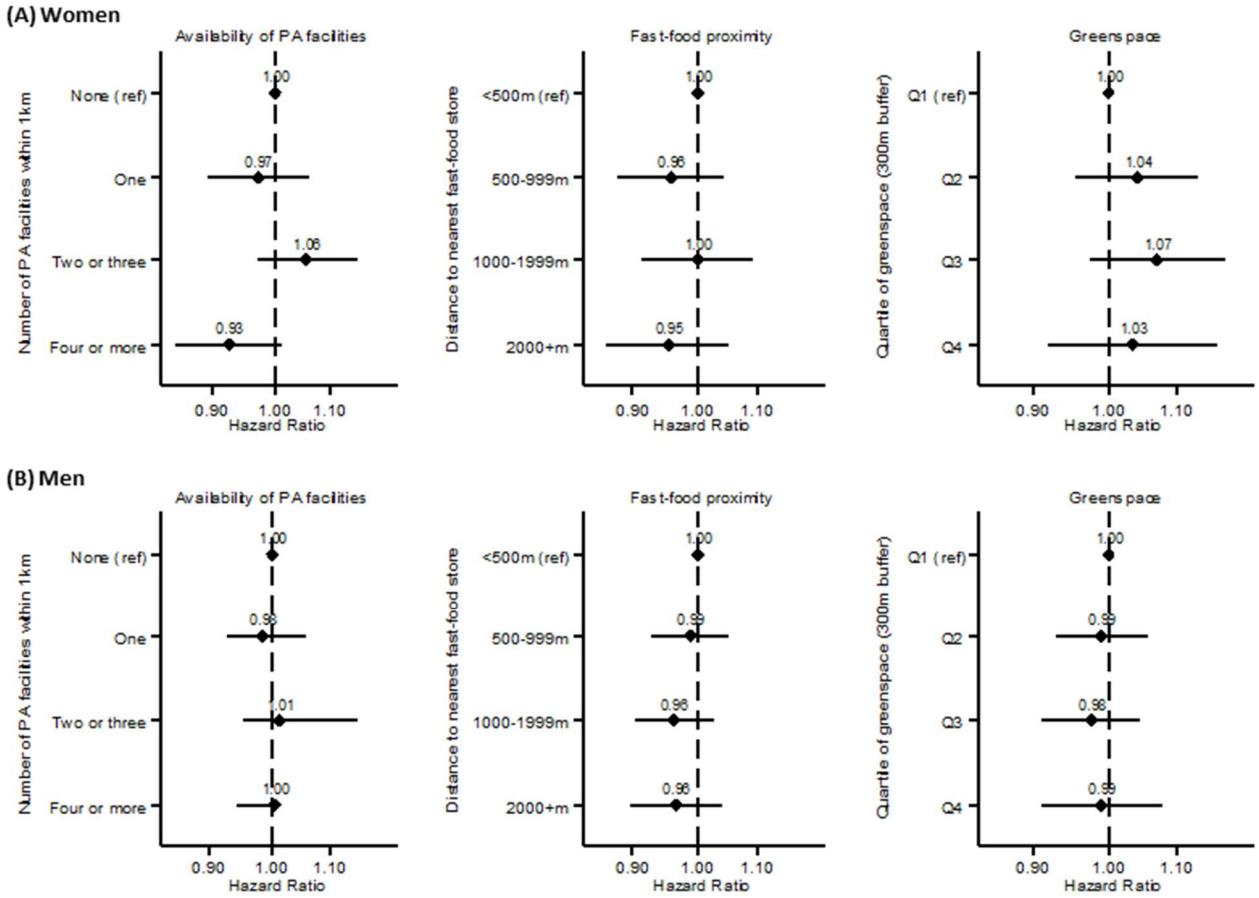
Supplementary Figure 11. Geographical heterogeneity in the association between distance to nearest fast-food/takeaway store and BMI, from separate regression models of 122 Local Authority Districts of England represented in UK Biobank

Notes: Significance refers to arbitrary threshold of P<0.05. Grey areas in the map were not included in UK Biobank, or were non-urban, or had <200 study participants in the LAD.

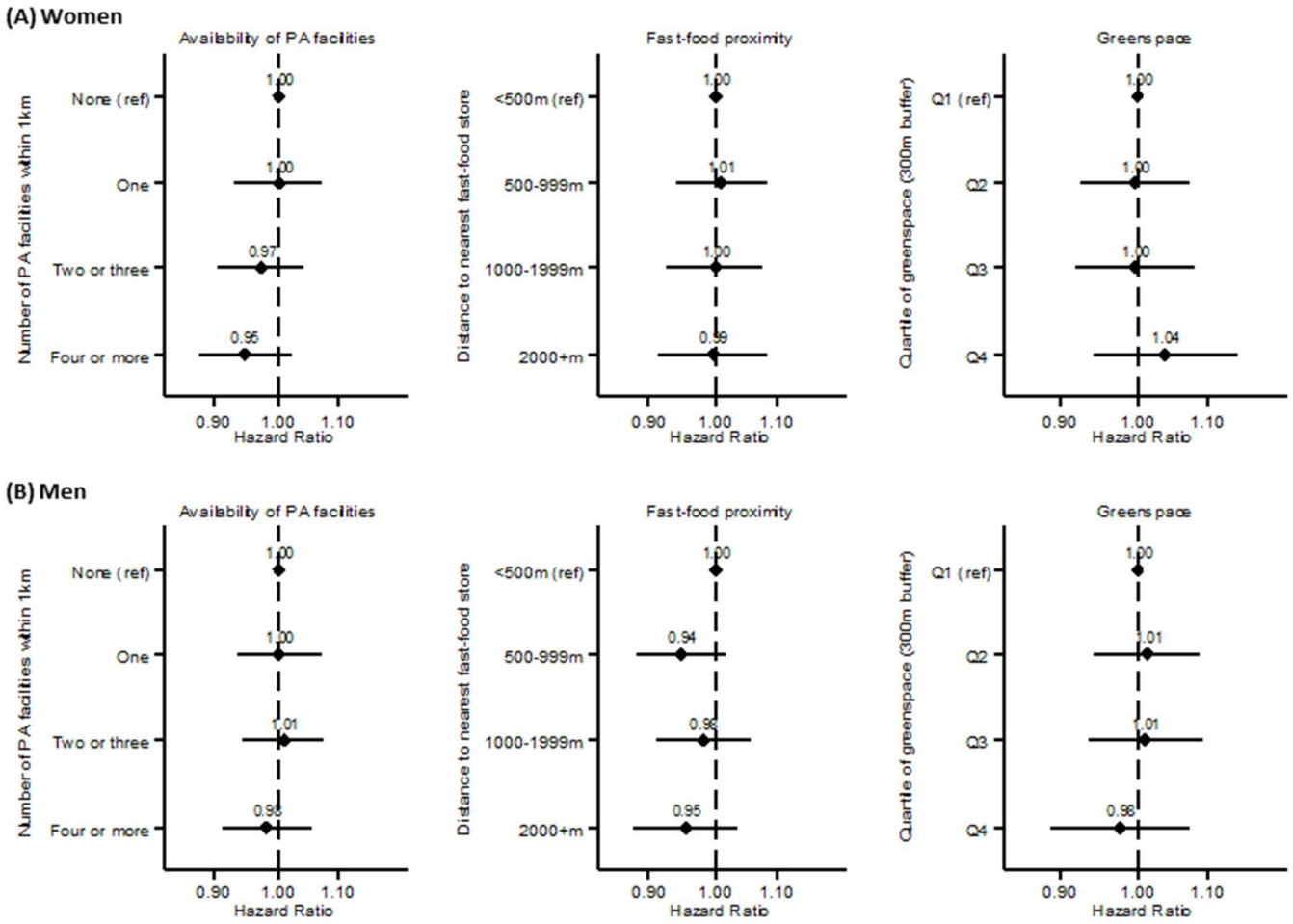
Appendix Five.

SUPPLEMENTARY MATERIAL: Research Paper 5

1. Primary analysis stratified by sex



Supplementary Figure 12. Hazard ratios for associations between neighbourhood characteristics and CVD-related hospital admissions, stratified by sex



Supplementary Figure 13. Hazard ratios for associations between neighbourhood characteristics and cancer-related hospital admissions, stratified by sex

Supplementary Table 8. Modification of the association between neighbourhood availability of PA facilities and hospital admissions due to CVD, by household income and area deprivation, stratified by sex

	Annual household income				*Area deprivation*			
	WOMEN		**MEN**		**WOMEN**		**MEN**	
	< £31,000	**At least £31,000**	**< £31,000**	**At least £31,000**	**More deprived**	**Less deprived**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.92 (0.82, 1.03) P=0.165	1.00 (ref)	1.01 (0.93, 1.09) P=0.814	1.00 (ref)	0.87 (0.76, 1.00) P=0.046	1.00 (ref)	0.89 (0.80, 0.98) P=0.015
One	0.96 (0.87, 1.06) P=0.435	0.91 (0.80, 1.05) P=0.196	1.03 (0.95, 1.11) P=0.539	0.94 (0.86, 1.04) P=0.222	0.93 (0.79, 1.10) P=0.415	0.86 (0.75, 0.99) P=0.038	0.93 (0.82, 1.05) P=0.235	0.89 (0.80, 0.98) P=0.020
2-3	1.05 (0.95, 1.15) P=0.353	0.98 (0.87, 1.12) P=0.809	1.03 (0.95, 1.11) P=0.449	1.00 (0.91, 1.09) P=0.968	1.09 (0.94, 1.26) P=0.261	0.91 (0.79, 1.04) P=0.160	0.93 (0.83, 1.04) P=0.179	0.93 (0.84, 1.02) P=0.138
4 or more	0.93 (0.84, 1.03) P=0.167	0.83 (0.73, 0.95) P=0.006	1.05 (0.97, 1.14) P=0.248	0.95 (0.87, 1.04) P=0.286	0.89 (0.77, 1.04) P=0.145	0.83 (0.72, 0.95) P=0.009	0.93 (0.83, 1.04) P=0.181	0.92 (0.83, 1.02) P=0.095
Stratum-specific HRs (4+ facilities vs 0)	0.94 (0.84, 1.05) P=0.245	0.88 (0.76, 1.04) P=0.129	1.04 (0.96, 1.14) P=0.327	0.94 (0.86, 1.04) P=0.227	0.92 (0.78, 1.09) P=0.334	0.94 (0.85, 1.05) P=0.308	0.89 (0.79, 1.01) P=0.064	1.04 (0.96, 1.12) P=0.315
Relative excess risk due to interaction (RERI)	-0.018 (-0.168, 0.132) P=0.813		-0.105 (-0.218, 0.009) P=0.070		0.061 (-0.096, 0.218) P=0.446		0.102 (-0.012, 0.217) P=0.081	

Supplementary Table 9. Modification of the association between fast-food proximity and hospital admissions due to CVD, by household income and area deprivation, stratified by sex

	Annual household income				*Area deprivation*			
	WOMEN		**MEN**		**WOMEN**		**MEN**	
	< £31,000	**At least £31,000**	**< £31,000**	**At least £31,000**	**More deprived**	**Less deprived**	**More deprived**	**Less deprived**
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.96 (0.83, 1.11) P=0.609	1.00 (ref)	0.90 (0.81, 0.99) P=0.039	1.00 (ref)	0.93 (0.81, 1.05) P=0.246	1.00 (ref)	1.01 (0.92, 1.11) P=0.851
500-999m	0.97 (0.88, 1.07) P=0.539	0.89 (0.78, 1.01) P=0.079	0.97 (0.90, 1.05) P=0.521	0.91 (0.83, 1.00) P=0.044	0.99 (0.88, 1.13) P=0.934	0.85 (0.76, 0.96) P=0.009	1.02 (0.93, 1.12) P=0.661	0.96 (0.88, 1.05) P=0.391
1000-1999m	1.03 (0.93, 1.14) P=0.584	0.90 (0.79, 1.02) P=0.111	0.93 (0.86, 1.01) P=0.087	0.91 (0.83, 1.00) P=0.044	0.99 (0.86, 1.14) P=0.889	0.91 (0.81, 1.03) P=0.120	0.96 (0.87, 1.07) P=0.479	0.95 (0.87, 1.03) P=0.225
At least 2000m	0.95 (0.85, 1.07) P=0.395	0.91 (0.79, 1.04) P=0.171	0.92 (0.85, 1.01) P=0.086	0.92 (0.84, 1.01) P=0.091	1.01 (0.86, 1.20) P=0.868	0.85 (0.74, 0.97) P=0.014	1.03 (0.91, 1.17) P=0.616	0.94 (0.85, 1.03) P=0.168
Stratum-specific HRs (≥2000m vs <500m)	0.93 (0.82, 1.05) P=0.226	0.99 (0.82, 1.18) P=0.880	0.92 (0.84, 1.02) P=0.107	1.02 (0.91, 1.14) P=0.722	0.98 (0.82, 1.18) P=0.857	0.92 (0.81, 1.04) P=0.201	1.06 (0.92, 1.22) P=0.394	0.92 (0.84, 1.00) P=0.057
Relative excess risk due to interaction (RERI)	-0.007 (-0.183, 0.170) P=0.942		0.098 (-0.017, 0.213) P=0.095		-0.092 (-0.290, 0.106) P=0.361		-0.107 (-0.259, 0.046) P=0.169	

Supplementary Table 10. Modification of the association between neighbourhood greenspace and hospital admissions due to CVD, by household income and area deprivation, stratified by sex

	Annual household income				*Area deprivation*			
	WOMEN		**MEN**		**WOMEN**		**MEN**	
	< £31,000	**At least £31,000**	**< £31,000**	**At least £31,000**	**More deprived**	**Less deprived**	**More deprived**	**Less deprived**
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.91 (0.79, 1.04) P=0.154	1.00 (ref)	0.91 (0.83, 1.00) P=0.046	1.00 (ref)	0.86 (0.75, 0.97) P=0.016	1.00 (ref)	1.05 (0.95, 1.15) P=0.339
Q2	1.03 (0.93, 1.14) P=0.608	0.97 (0.85, 1.11) P=0.702	0.98 (0.90, 1.06) P=0.598	0.92 (0.83, 1.01) P=0.067	1.01 (0.89, 1.14) P=0.874	0.90 (0.81, 1.00) P=0.057	1.03 (0.94, 1.13) P=0.576	0.97 (0.89, 1.05) P=0.389
Q3	1.06 (0.95, 1.18) P=0.315	0.99 (0.87, 1.14) P=0.941	0.95 (0.88, 1.04) P=0.278	0.92 (0.84, 1.01) P=0.083	1.04 (0.89, 1.21) P=0.621	0.92 (0.82, 1.02) P=0.116	1.05 (0.93, 1.17) P=0.446	0.94 (0.87, 1.02) P=0.154
Q4 (most greenspace)	1.04 (0.91, 1.18) P=0.587	0.93 (0.80, 1.08) P=0.330	0.95 (0.86, 1.05) P=0.276	0.95 (0.86, 1.05) P=0.331	0.97 (0.79, 1.19) P=0.759	0.89 (0.78, 1.01) P=0.070	0.98 (0.84, 1.15) P=0.847	0.97 (0.88, 1.06) P=0.497
Stratum-specific HRs (Q4 vs Q1)	1.03 (0.89, 1.19) P=0.691	1.03 (0.84, 1.27) P=0.758	0.97 (0.86, 1.08) P=0.565	1.01 (0.89, 1.15) P=0.859	0.86 (0.67, 1.11) P=0.244	1.06 (0.91, 1.22) P=0.456	1.02 (0.84, 0.23) P=0.878	0.91 (0.82. 1.01) P=0.077
Relative excess risk due to interaction (RERI)	-0.015 (-0.186, 0.155) P=0.861		0.096 (-0.017, 0.208) P=0.095		0.063 (-0.155, 0.280) P=0.573		-0.062 (-0.233, 0.109) P=0.475	

Supplementary Table 11. Modification of the association between neighbourhood availability of PA facilities and hospital admissions due to cancer, by household income and area deprivation, stratified by sex

	Annual household income				Area deprivation			
	WOMEN		MEN		WOMEN		MEN	
	< £31,000	At least £31,000	< £31,000	At least £31,000	More deprived	Less deprived	More deprived	Less deprived
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.89 (0.81, 0.98) P=0.018	1.00 (ref)	1.03 (0.94, 1.13) P=0.482	1.00 (ref)	1.08 (0.96, 1.21) P=0.217	1.00 (ref)	0.97 (0.87, 1.09) P=0.602
One	0.95 (0.87, 1.04) P=0.236	0.96 (0.87, 1.07) P=0.501	1.02 (0.93, 1.11) P=0.703	1.00 (0.91, 1.11) P=0.932	1.09 (0.94, 1.26) P=0.251	1.05 (0.93, 1.19) P=0.407	1.06 (0.92, 1.22) P=0.421	0.95 (0.85, 1.07) P=0.415
2-3	0.98 (0.90, 1.07) P=0.668	0.85 (0.77, 0.94) P=0.002	1.02 (0.93, 1.11) P=0.700	1.02 (0.93, 1.13) P=0.652	1.13 (0.99, 1.29) P=0.081	0.99 (0.87, 1.12) P=0.876	1.00 (0.88, 1.13) P=0.951	0.99 (0.88, 1.11) P=0.865
4 or more	0.96 (0.88, 1.05) P=0.401	0.82 (0.74, 0.91) P=0.000	1.01 (0.93, 1.11) P=0.742	0.96 (0.87, 1.06) P=0.405	1.00 (0.87, 1.14) P=0.959	1.02 (0.90, 1.16) P=0.758	1.02 (0.90, 1.16) P=0.745	0.94 (0.83, 1.05) P=0.270
Stratum-specific HRs (4+ facilities vs 0)	0.94 (0.85, 1.03) P=0.181	0.96 (0.85, 1.08) P=0.532	1.01 (0.92, 1.11) P=0.864	0.94 (0.84, 1.05) P=0.243	1.03 (0.89, 1.19) P=0.711	0.94 (0.86, 1.02) P=0.152	1.02 (0.89, 1.17) P=0.808	0.96 (0.88, 1.05) P=0.391
Relative excess risk due to interaction (RERI)	-0.034 (-0.154, 0.087) P=0.584		-0.089 (-0.216, 0.039) P=0.172		-0.054 (-0.213, 0.104) P=0.502		-0.056 (-0.203, 0.091) P=0.454	

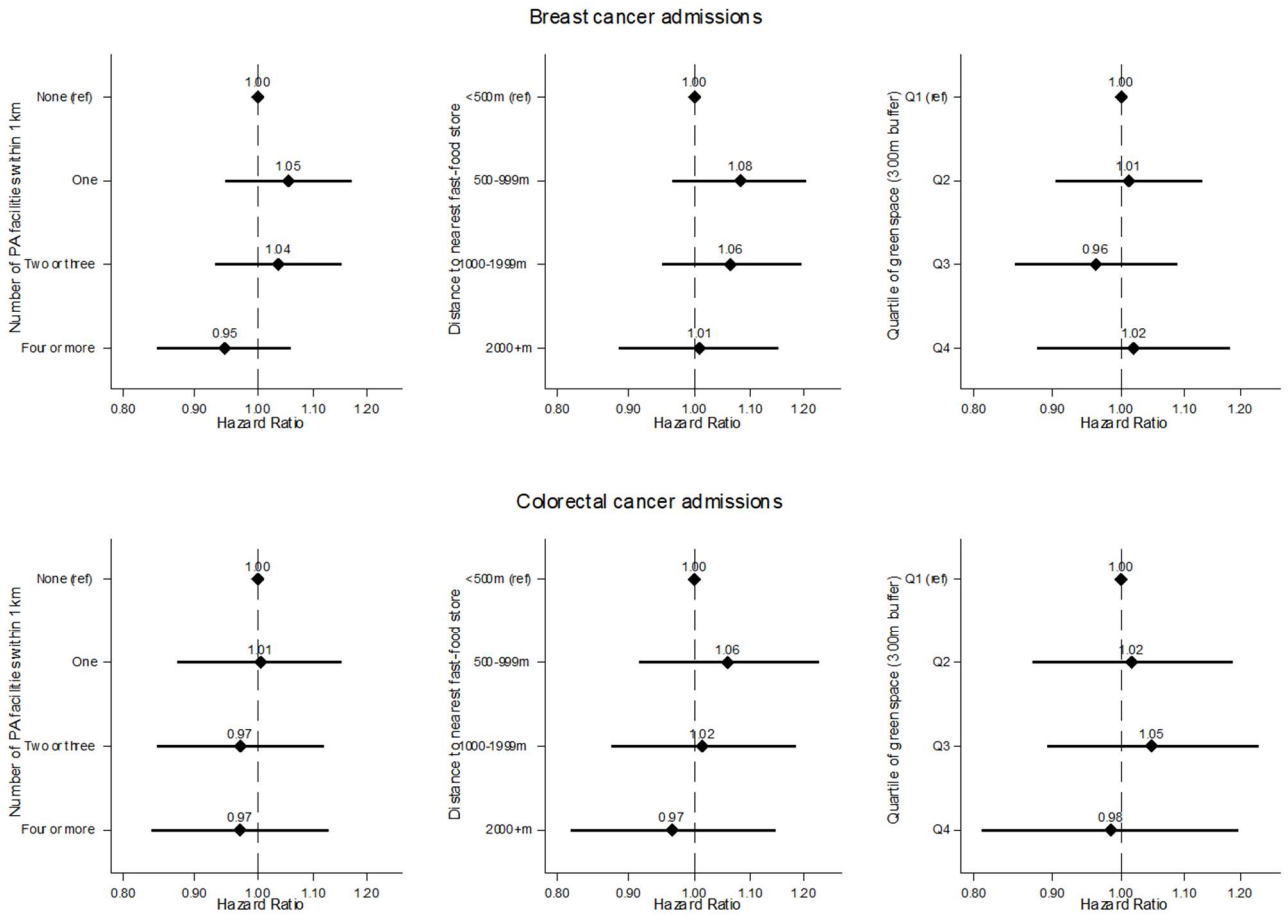
Supplementary Table 12. Modification of the association between fast-food proximity and hospital admissions due to cancer, by household income and area deprivation, stratified by sex

	Annual household income				Area deprivation			
	WOMEN		MEN		WOMEN		MEN	
	< £31,000	At least £31,000	< £31,000	At least £31,000	More deprived	Less deprived	More deprived	Less deprived
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.78 (0.69, 0.88) P=0.000	1.00 (ref)	1.00 (0.89, 1.12) P=0.971	1.00 (ref)	0.92 (0.82, 1.03) P=0.138	1.00 (ref)	0.91 (0.82, 1.01) P=0.088
500-999m	0.94 (0.86, 1.03) P=0.174	0.90 (0.81, 1.00) P=0.050	0.94 (0.86, 1.03) P=0.160	0.95 (0.86, 1.06) P=0.349	0.95 (0.85, 1.06) P=0.368	0.97 (0.87, 1.07) P=0.527	0.85 (0.77, 0.95) P=0.005	0.91 (0.83, 1.00) P=0.058
1000-1999m	0.97 (0.88, 1.06) P=0.461	0.84 (0.76, 0.94) P=0.002	0.99 (0.90, 1.08) P=0.762	0.97 (0.87, 1.07) P=0.547	0.95 (0.84, 1.08) P=0.452	0.95 (0.86, 1.05) P=0.314	1.03 (0.92, 1.15) P=0.649	0.89 (0.80, 0.98) P=0.015
At least 2000m	0.93 (0.84, 1.03) P=0.141	0.88 (0.79, 0.99) P=0.028	0.96 (0.87, 1.06) P=0.394	0.95 (0.85, 1.05) P=0.311	0.94 (0.81, 1.09) P=0.420	0.95 (0.85, 1.06) P=0.320	0.92 (0.79, 1.06) P=0.235	0.89 (0.80, 0.98) P=0.024
Stratum-specific HRs (≥2000m vs <500m)	0.95 (0.85, 1.06) P=0.364	1.08 (0.93, 1.24) P=0.314	0.99 (0.89, 1.10) P=0.847	0.90 (0.79, 1.02) P=0.101	0.93 (0.79, 1.10) P=0.403	1.03 (0.93, 1.15) P=0.521	0.91 (0.78, 1.07) P=0.264	0.97 (0.88, 1.07) P=0.515
Relative excess risk due to interaction (RERI)	0.174 (0.047, 0.301) P=0.007		-0.010 (-0.149, 0.129) P=0.886		0.087 (-0.075, 0.249) P=0.294		0.059 (-0.095, 0.213) P=0.451	

Supplementary Table 13. Modification of the association between neighbourhood greenspace and hospital admissions due to cancer, by household income and area deprivation, stratified by sex

	Annual household income				Area deprivation			
	WOMEN		MEN		WOMEN		MEN	
	< £31,000	At least £31,000	< £31,000	At least £31,000	More deprived	Less deprived	More deprived	Less deprived
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.79 (0.71, 0.88) P=0.000	1.00 (ref)	0.93 (0.84, 1.04) P=0.216	1.00 (ref)	0.97 (0.87, 1.08) P=0.538	1.00 (ref)	0.90 (0.81, 1.00) P=0.049
Q2	0.94 (0.86, 1.03) P=0.202	0.88 (0.79, 0.98) P=0.020	1.02 (0.93, 1.11) P=0.685	0.94 (0.85, 1.05) P=0.258	1.02 (0.91, 1.13) P=0.786	0.96 (0.87, 1.05) P=0.347	0.96 (0.87, 1.07) P=0.480	0.94 (0.86, 1.03) P=0.191
Q3	0.97 (0.89, 1.07) P=0.591	0.84 (0.75, 0.94) P=0.002	0.95 (0.86, 1.04) P=0.281	1.04 (0.93, 1.15) P=0.495	0.98 (0.86, 1.13) P=0.795	0.97 (0.88, 1.06) P=0.498	1.01 (0.89, 1.15) P=0.845	0.92 (0.84, 1.01) P=0.078
Q4 (most greenspace)	0.95 (0.85, 1.06) P=0.330	0.96 (0.85, 1.08) P=0.461	0.95 (0.85, 1.06) P=0.348	0.96 (0.85, 1.08) P=0.492	0.87 (0.72, 1.06) P=0.166	1.02 (0.92, 1.14) P=0.681	0.85 (0.71, 1.02) P=0.078	0.91 (0.82, 1.01) P=0.066
Stratum-specific HRs (Q4 vs Q1)	0.98 (0.86, 1.11) P=0.756	1.11 (0.95, 1.30) P=0.180	0.96 (0.84, 1.09) P=0.503	1.00 (0.87, 1.16) P=0.985	0.83 (0.66, 1.04) P=0.112	1.08 (0.96, 1.21) P=0.205	0.85 (0.68, 1.05) P=0.131	1.01 (0.90, 1.13) P=0.893
Relative excess risk due to interaction (RERI)	0.220 (0.097, 0.344) P<0.001		0.077 (-0.052, 0.207) P=0.243		0.182 (-0.002, 0.367) P=0.053		0.158 (-0.013, 0.330) P=0.071	

2. Secondary outcomes



Supplementary Figure 14. Hazard ratios for associations between neighbourhood characteristics and hospital admissions for breast cancer (women only) and colorectal cancer (all)

Supplementary Table 14. Modification of the association between built environment variables and hospital admissions due to breast cancer, by household income and area deprivation (women only)

Breast cancer-related admissions	*Annual household income*		*Area deprivation*	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	1.00 (0.87, 1.15) P=0.977	1.00 (ref)	1.15 (0.95, 1.38) P=0.142
One	1.03 (0.89, 1.19) P=0.683	1.08 (0.92, 1.26) P=0.336	1.16 (0.92, 1.46) P=0.196	1.18 (0.97, 1.43) P=0.090
2-3	1.07 (0.94, 1.23) P=0.313	0.99 (0.85, 1.15) P=0.889	1.17 (0.95, 1.45) P=0.137	1.14 (0.94, 1.38) P=0.183
4 or more	1.02 (0.88, 1.18) P=0.778	0.86 (0.73, 1.00) P=0.056	0.95 (0.77, 1.18) P=0.647	1.10 (0.91, 1.34) P=0.315
Stratum-specific HRs (4+ facilities vs 0)	1.03 (0.88, 1.20) P=0.731	0.86 (0.72, 1.01) P=0.071	1.03 (0.81, 1.29) P=0.827	0.94 (0.82, 1.07) P=0.331
Relative excess risk due to interaction (RERI)	-0.161 (-0.358, 0.036) P=0.109		0.004 (-0.239, 0.247) P=0.975	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.77 (0.64, 0.93) P=0.006	1.00 (ref)	0.98 (0.82, 1.16) P=0.801
500-999m	0.95 (0.82, 1.09) P=0.447	1.01 (0.86, 1.17) P=0.942	0.96 (0.81, 1.15) P=0.690	1.14 (0.97, 1.33) P=0.113
1000-1999m	0.98 (0.84, 1.13) P=0.736	0.94 (0.80, 1.10) P=0.449	1.07 (0.89, 1.29) P=0.468	1.07 (0.91, 1.26) P=0.405
At least 2000m	0.94 (0.80, 1.10) P=0.437	0.88 (0.74, 1.04) P=0.126	0.91 (0.72, 1.16) P=0.452	1.04 (0.88, 1.24) P=0.645
Stratum-specific HRs (≥2000m vs <500m)	0.95 (0.80, 1.13) P=0.548	1.12 (0.91, 1.36) P=0.291	0.86 (0.66, 1.12) P=0.254	1.08 (0.92, 1.26) P=0.334
Relative excess risk due to interaction (RERI)	0.162 (-0.028, 0.353) P=0.094		0.151 (-0.103, 0.405) P=0.243	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.82 (0.70, 0.97) P=0.019	1.00 (ref)	1.00 (0.86, 1.18) P=0.954
Q2	0.93 (0.80, 1.07) P=0.279	0.95 (0.81, 1.11) P=0.525	1.00 (0.84, 1.18) P=0.982	1.06 (0.92, 1.22) P=0.432
Q3	0.98 (0.84, 1.14) P=0.784	0.79 (0.67, 0.94) P=0.007	0.95 (0.76, 1.18) P=0.629	1.02 (0.88, 1.17) P=0.813
Q4 (most greenspace)	0.87 (0.73, 1.04) P=0.134	1.00 (0.84, 1.20) P=0.958	0.78 (0.57, 1.07) P=0.123	1.10 (0.94, 1.30) P=0.240
Stratum-specific HRs (Q4 vs Q1)	0.84 (0.68, 1.03) P=0.097	1.23 (0.99, 1.53) P=0.060	0.69 (0.47, 0.99) P=0.044	1.14 (0.96, 1.35) P=0.137
Relative excess risk due to interaction (RERI)	0.307 (0.126, 0.488) P=0.001		0.316 (0.041, 0.591) P=0.024	

Supplementary Table 15. Association between neighbourhood characteristics and breast cancer-related hospital admissions, stratified by household income and area deprivation in combination

Breast cancer-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
One	1.22 (0.92, 1.61) P=0.165	1.13 (0.74, 1.73) P=0.581	0.97 (0.81, 1.15) P=0.698	1.08 (0.91, 1.27) P=0.385
2-3	1.29 (0.99, 1.68) P=0.055	1.07 (0.71, 1.60) P=0.747	0.97 (0.82, 1.16) P=0.757	0.98 (0.82, 1.16) P=0.794
4 or more	1.11 (0.84, 1.48) P=0.450	0.85 (0.56, 1.29) P=0.444	1.01 (0.84, 1.22) P=0.912	0.87 (0.72, 1.05) P=0.158
Fast-food proximity				
Closer than 500m	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
500-999m	0.87 (0.70, 1.09) P=0.234	1.11 (0.82, 1.51) P=0.511	1.00 (0.82, 1.22) P=0.985	1.40 (1.12, 1.74) P=0.003
1000-1999m	1.00 (0.78, 1.27) P=0.977	1.05 (0.73, 1.50) P=0.799	0.99 (0.81, 1.21) P=0.955	1.27 (1.02, 1.59) P=0.031
At least 2000m	0.85 (0.62, 1.17) P=0.321	0.87 (0.53, 1.43) P=0.575	0.99 (0.80, 1.22) P=0.902	1.22 (0.96, 1.54) P=0.104
Greenspace				
Q1 (least greenspace)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Q2	0.93 (0.75, 1.15) P=0.489	0.99 (0.72, 1.37) P=0.973	0.95 (0.77, 1.17) P=0.636	1.13 (0.91, 1.41) P=0.252
Q3	0.94 (0.72, 1.23) P=0.639	0.64 (0.40, 1.03) P=0.068	1.05 (0.85, 1.30) P=0.629	0.99 (0.79, 1.23) P=0.898
Q4 (most greenspace)	0.61 (0.38, 0.97) P=0.036	0.80 (0.43, 1.50) P=0.486	0.92 (0.72, 1.19) P=0.533	1.35 (1.05, 1.73) P=0.017

Supplementary Table 16. Modification of the association between built environment variables and hospital admissions due to colorectal cancer, by household income and area deprivation

Colorectal cancer-related admissions	*Annual household income*		*Area deprivation*	
	Less than £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.85 (0.71, 1.02) P=0.077	1.00 (ref)	1.31 (1.02, 1.70) P=0.035
One	0.95 (0.80, 1.13) P=0.566	0.93 (0.75, 1.14) P=0.477	1.34 (0.98, 1.83) P=0.068	1.24 (0.95, 1.62) P=0.111
2-3	0.92 (0.77, 1.10) P=0.352	0.89 (0.73, 1.09) P=0.274	1.18 (0.88, 1.59) P=0.268	1.23 (0.95, 1.60) P=0.118
4 or more	0.92 (0.77, 1.10) P=0.381	0.89 (0.72, 1.09) P=0.243	1.26 (0.95, 1.67) P=0.112	1.18 (0.91, 1.55) P=0.216
Stratum-specific HRs (4+ facilities vs 0)	0.87 (0.72, 1.06) P=0.169	1.13 (0.89, 1.42) P=0.313	1.33 (0.97, 1.81) P=0.073	0.87 (0.73, 1.04) P=0.123
Relative excess risk due to interaction (RERI)	0.115 (-0.114, 0.345) P=0.325		-0.389 (-0.811, 0.033) P=0.071	
	Less than £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.91 (0.71, 1.16) P=0.431	1.00 (ref)	0.99 (0.78, 1.25) P=0.924
500-999m	1.12 (0.93, 1.34) P=0.225	0.88 (0.71, 1.09) P=0.248	0.94 (0.74, 1.19) P=0.582	1.12 (0.91, 1.37) P=0.296
1000-1999m	1.03 (0.85, 1.25) P=0.744	0.91 (0.73, 1.13) P=0.405	1.12 (0.87, 1.43) P=0.368	0.99 (0.80, 1.22) P=0.914
At least 2000m	0.87 (0.71, 1.07) P=0.195	1.00 (0.80, 1.25) P=0.984	0.87 (0.63, 1.21) P=0.414	0.99 (0.79, 1.24) P=0.917
Stratum-specific HRs (≥2000m vs <500m)	0.89 (0.71, 1.11) P=0.290	1.06 (0.81, 1.40) P=0.650	0.87 (0.61, 1.24) P=0.440	1.00 (0.82, 1.23) P=0.967
Relative excess risk due to interaction (RERI)	0.220 (-0.048, 0.488) P=0.107		0.125 (-0.208, 0.457) P=0.462	
	Less than £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.87 (0.69, 1.09) P=0.214	1.00 (ref)	0.96 (0.77, 1.20) P=0.745
Q2	1.02 (0.85, 1.23) P=0.805	0.88 (0.70, 1.10) P=0.248	0.98 (0.78, 1.23) P=0.844	1.02 (0.85, 1.23) P=0.825
Q3	0.99 (0.81, 1.20) P=0.911	1.00 (0.81, 1.24) P=0.991	1.04 (0.79, 1.38) P=0.760	1.05 (0.87, 1.26) P=0.643
Q4 (most greenspace)	0.96 (0.77, 1.20) P=0.717	0.89 (0.70, 1.14) P=0.367	0.77 (0.51, 1.15) P=0.203	1.01 (0.81, 1.25) P=0.962
Stratum-specific HRs (Q4 vs Q1)	1.04 (0.80, 1.34) P=0.779	0.92 (0.68, 1.26) P=0.619	0.69 (0.43, 1.12) P=0.135	1.08 (0.86, 1.37) P=0.494
Relative excess risk due to interaction (RERI)	0.069 (-0.189, 0.327) P=0.601		0.272 (-0.083, 0.627) P=0.133	

Supplementary Table 17. Association between neighbourhood characteristics and colorectal cancer-related hospital admissions, stratified by household income and area deprivation in combination

Colorectal cancer-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
	HR (95% CI)	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Number of PA facilities				
None	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
One	1.37 (0.95, 1.98) P=0.090	1.33 (0.71, 2.48) P=0.373	0.83 (0.67, 1.02) P=0.077	1.10 (0.87, 1.39) P=0.421
2-3	1.28 (0.90, 1.83) P=0.173	1.11 (0.61, 2.03) P=0.724	0.80 (0.65, 0.99) P=0.043	1.12 (0.88, 1.42) P=0.367
4 or more	1.38 (0.96, 1.99) P=0.084	1.21 (0.67, 2.19) P=0.533	0.72 (0.56, 0.91) P=0.006	1.11 (0.86, 1.44) P=0.414
Fast-food proximity				
Closer than 500m	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
500-999m	1.01 (0.76, 1.34) P=0.949	0.83 (0.53, 1.31) P=0.427	1.22 (0.96, 1.56) P=0.111	1.00 (0.74, 1.35) P=0.999
1000-1999m	1.10 (0.81, 1.50) P=0.549	1.25 (0.77, 2.03) P=0.359	1.05 (0.81, 1.35) P=0.713	0.93 (0.69, 1.26) P=0.638
At least 2000m	0.84 (0.55, 1.27) P=0.400	0.99 (0.51, 1.92) P=0.984	0.92 (0.69, 1.21) P=0.539	1.09 (0.80, 1.49) P=0.592
Greenspace				
Q1 (least greenspace)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Q2	1.03 (0.79, 1.36) P=0.807	0.68 (0.41, 1.12) P=0.132	1.08 (0.82, 1.40) P=0.593	1.07 (0.77, 1.47) P=0.697
Q3	1.05 (0.74, 1.48) P=0.797	0.85 (0.47, 1.53) P=0.591	1.06 (0.81, 1.40) P=0.659	1.17 (0.85, 1.61) P=0.330
Q4 (most greenspace)	0.68 (0.37, 1.23) P=0.203	0.67 (0.28, 1.59) P=0.366	1.14 (0.83, 1.56) P=0.42	1.05 (0.73, 1.50) P=0.802

3. Sensitivity analyses: Restricting follow-up time to January 2012 onwards

Supplementary Table 18. Hospital admissions by household income and area deprivation (follow-up time restricted to January 2012 onwards)

	CVD		Cancer	
	N	Admissions (%)	N	Admissions (%)
Total	330045	7698 (2.3)	320812	8168 (2.5)
Household income (annual pre-tax)				
<£31,000	154130	4497 (2.9)	151885	4728 (3.1)
£31,000 or more	175915	3201 (1.8)	168927	3440 (2.0)
Area deprivation				
More deprived	95164	2342 (2.5)	94194	2419 (2.6)
Less deprived	234881	5356 (2.3)	226618	5749 (2.5)

Supplementary Table 19. Modification of the associations between built environment variables and hospital admissions due to CVD, by household income and area deprivation (follow-up time restricted to January 2012 onwards)

CVD-related admissions	*Annual household income*		*Area deprivation*	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.99 (0.92, 1.10) P=0.934	1.00 (ref)	0.89 (0.80, 0.98) P=0.024
One	1.00 (0.91, 1.08) P=0.852	0.94 (0.88, 1.08) P=0.577	0.89 (0.77, 1.01) P=0.077	0.90 (0.81, 1.01) P=0.061
2-3	1.04 (0.96, 1.13) P=0.290	0.99 (0.92, 1.12) P=0.727	1.01 (0.90, 1.14) P=0.852	0.92 (0.82, 1.02) P=0.119
4 or more	1.00 (0.95, 1.13) P=0.393	0.92 (0.84, 1.02) P=0.133	0.94 (0.84, 1.06) P=0.326	0.90 (0.80, 1.00) P=0.058
Stratum-specific HRs (4+ facilities vs 0)	1.03 (0.95, 1.13) P=0.466	0.92 (0.83, 1.02) P=0.130	0.94 (0.83, 1.07) P=0.354	1.01 (0.93, 1.10) P=0.772
Relative excess risk due to interaction (RERI)	-0.113 (-0.236, 0.010) P=0.072		0.070 (-0.056, 0.197) P=0.274	
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.90 (0.80, 1.00) P=0.059	1.00 (ref)	0.98 (0.88, 1.08) P=0.667
500-999m	0.94 (0.86, 1.02) P=0.141	0.91 (0.82, 1.00) P=0.051	1.01 (0.91, 1.12) P=0.864	0.90 (0.82, 1.00) P=0.039
1000-1999m	0.95 (0.87, 1.04) P=0.247	0.93 (0.85, 1.03) P=0.173	0.98 (0.88, 1.10) P=0.766	0.94 (0.86, 1.03) P=0.207
At least 2000m	0.94 (0.85, 1.03) P=0.182	0.92 (0.83, 1.02) P=0.132	1.00 (0.87, 1.15) P=0.977	0.92 (0.83, 1.02) P=0.118
Stratum-specific HRs (≥2000m vs <500m)	0.94 (0.85, 1.04) P=0.243	1.02 (0.90, 1.16) P=0.725	0.99 (0.85, 1.15) P=0.874	0.95 (0.86, 1.04) P=0.263
Relative excess risk due to interaction (RERI)	0.087 (-0.040, 0.213) P=0.179		-0.058 (-0.217, 0.101) P=0.476	
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.91 (0.83, 1.01) P=0.090	1.00 (ref)	0.92 (0.84, 1.02) P=0.118
Q2	1.00 (0.92, 1.08) P=0.932	0.94 (0.85, 1.04) P=0.209	0.97 (0.88, 1.08) P=0.599	0.94 (0.86, 1.02) P=0.134
Q3	0.97 (0.89, 1.07) P=0.576	0.99 (0.90, 1.10) P=0.892	1.03 (0.91, 1.17) P=0.610	0.93 (0.85, 1.01) P=0.088
Q4 (most greenspace)	0.98 (0.88, 1.09) P=0.717	0.95 (0.84, 1.06) P=0.338	0.94 (0.80, 1.12) P=0.499	0.92 (0.83, 1.02) P=0.120
Stratum-specific HRs (Q4 vs Q1)	0.99 (0.88, 1.12) P=0.908	1.02 (0.89, 1.18) P=0.738	0.88 (0.72, 1.08) P=0.228	1.01 (0.90, 1.12) P=0.925
Relative excess risk due to interaction (RERI)	0.050 (-0.076, 0.176) P=0.435		0.056 (-0.119, 0.231) P=0.530	

Supplementary Table 20. Modification of the associations between neighbourhood environment variables and hospital admissions due to cancer, by household income and area deprivation (follow-up time restricted to January 2012 onwards)

Cancer-related admissions	*Annual household income*		*Area deprivation*	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.98 (0.90, 1.07) P=0.669	1.00 (ref)	1.01 (0.91, 1.12) P=0.846
One	0.96 (0.88, 1.04) P=0.317	0.94 (0.85, 1.04) P=0.212	1.07 (0.94, 1.22) P=0.276	0.93 (0.84, 1.04) P=0.220
2-3	0.99 (0.91, 1.07) P=0.781	0.90 (0.82, 0.98) P=0.022	1.00 (0.89, 1.13) P=0.971	0.96 (0.86, 1.07) P=0.457
4 or more	0.96 (0.88, 1.04) P=0.328	0.89 (0.81, 0.98) P=0.014	0.96 (0.86, 1.08) P=0.538	0.95 (0.85, 1.06) P=0.385
Stratum-specific HRs (4+ facilities vs 0)	0.93 (0.86, 1.02) P=0.131	0.94 (0.84, 1.04) P=0.218	0.94 (0.83, 1.07) P=0.346	0.95 (0.87, 1.03) P=0.181
Relative excess risk due to interaction (RERI)	-0.053 (-0.166, 0.061) P=0.362		-0.022 (-0.157, 0.113) P=0.751	
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.86 (0.77, 0.96) P=0.008	1.00 (ref)	0.93 (0.84, 1.03) P=0.188
500-999m	0.92 (0.85, 1.00) P=0.053	0.91 (0.83, 1.01) P=0.065	0.91 (0.82, 1.01) P=0.077	0.94 (0.86, 1.03) P=0.170
1000-1999m	0.99 (0.91, 1.07) P=0.760	0.91 (0.83, 1.01) P=0.065	1.05 (0.95, 1.17) P=0.330	0.93 (0.85, 1.02) P=0.133
At least 2000m	0.94 (0.85, 1.03) P=0.163	0.93 (0.84, 1.02) P=0.132	0.91 (0.79, 1.05) P=0.184	0.94 (0.85, 1.04) P=0.236
Stratum-specific HRs (≥2000m vs <500m)	0.97 (0.88, 1.07) P=0.576	1.01 (0.89. 1.14) P=0.860	0.93 (0.80, 1.08) P=0.315	1.00 (0.91, 1.10) P=0.946
Relative excess risk due to interaction (RERI)	0.125 (0.004, 0.245) P=0.042		0.096 (-0.049, 0.242) P=0.196	
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.87 (0.79, 0.96) P=0.006	1.00 (ref)	0.94 (0.85, 1.04) P=0.218
Q2	0.98 (0.91, 1.07) P=0.668	0.87 (0.79, 0.97) P=0.008	0.98 (0.89, 1.09) P=0.740	0.93 (0.86, 1.01) P=0.106
Q3	0.93 (0.85, 1.02) P=0.109	0.94 (0.85, 1.04) P=0.244	0.96 (0.85, 1.09) P=0.567	0.93 (0.86, 1.01) P=0.105
Q4 (most greenspace)	0.93 (0.84, 1.04) P=0.199	0.94 (0.85, 1.05) P=0.300	0.86 (0.72, 1.02) P=0.081	0.95 (0.86, 1.05) P=0.306
Stratum-specific HRs (Q4 vs Q1)	0.97 (0.86, 1.09) P=0.574	1.02 (0.89, 1.17) P=0.768	0.87 (0.70, 1.06) P=0.166	1.01 (0.91, 1.23) P=0.789
Relative excess risk due to interaction (RERI)	0.139 (0.023, 0.255) P=0.019		0.150 (-0.014, 0.314) P=0.072	

Supplementary Table 21. Association between neighbourhood characteristics and CVD-related hospital admissions, stratified by household income and area deprivation in combination (follow-up time restricted to January 2012 onwards)

CVD-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None (ref)	1.00	1.00	1.00	1.00
One	0.95 (0.81, 1.11) P=0.524	0.75 (0.57, 0.97) P=0.030	1.01 (0.91, 1.13) P=0.789	1.01 (0.90, 1.13) P=0.847
2-3	1.08 (0.93, 1.25) P=0.338	0.85 (0.68, 1.07) P=0.174	1.03 (0.92, 1.14) P=0.627	1.03 (0.92, 1.15) P=0.611
4 or more	1.04 (0.89, 1.22) P=0.593	0.73 (0.57, 0.92) P=0.008	1.03 (0.92, 1.16) P=0.572	0.97 (0.86, 1.10) P=0.659
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m (ref)	1.00	1.00	1.00	1.00
500-999m	1.01 (0.89, 1.14) P=0.894	1.03 (0.85, 1.24) P=0.795	0.89 (0.79, 1.00) P=0.053	0.99 (0.86, 1.14) P=0.864
1000-1999m	0.95 (0.83, 1.09) P=0.483	1.08 (0.87, 1.34) P=0.493	0.95 (0.84, 1.07) P=0.401	1.00 (0.87, 1.15) P=0.977
At least 2000m	0.93 (0.78, 1.11) P=0.425	1.15 (0.87, 1.52) P=0.334	0.94 (0.82, 1.06) P=0.311	0.99 (0.85, 1.15) P=0.866
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace) (ref)	1.00	1.00	1.00	1.00
Q2	0.94 (0.83, 1.06) P=0.313	1.08 (0.89, 1.31) P=0.425	1.04 (0.92, 1.18) P=0.535	1.00 (0.86, 1.16) P=0.993
Q3	1.03 (0.88, 1.20) P=0.695	0.98 (0.76, 1.28) P=0.901	0.97 (0.85, 1.11) P=0.649	1.10 (0.95, 1.28) P=0.211
Q4 (most greenspace)	0.92 (0.72, 1.17) P=0.494	0.81 (0.54, 1.20) P=0.289	1.02 (0.87, 1.19) P=0.804	1.05 (0.89, 1.25) P=0.543

Supplementary Table 22. Association between neighbourhood characteristics and cancer-related hospital admissions, stratified by household income and area deprivation in combination (follow-up time restricted to January 2012 onwards)

Cancer-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None (ref)	1.00	1.00	1.00	1.00
One	1.01 (0.87, 1.18) P=0.876	1.18 (0.91, 1.53) P=0.203	0.92 (0.83, 1.02) P=0.106	0.94 (0.84, 1.04) P=0.232
2-3	1.00 (0.86, 1.15) P=0.964	0.95 (0.74, 1.22) P=0.681	0.96 (0.87, 1.07) P=0.488	0.94 (0.84, 1.05) P=0.263
4 or more	0.94 (0.80, 1.09) P=0.390	0.95 (0.75, 1.22) P=0.702	0.93 (0.83, 1.04) P=0.216	0.95 (0.85, 1.07) P=0.425
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m (ref)	1.00	1.00	1.00	1.00
500-999m	0.93 (0.82, 1.05) P=0.221	0.90 (0.74, 1.08) P=0.267	0.94 (0.84, 1.06) P=0.308	1.10 (0.96, 1.27) P=0.153
1000-1999m	1.05 (0.92, 1.20) P=0.504	1.09 (0.88, 1.34) P=0.432	1.00 (0.89, 1.12) P=0.995	1.01 (0.88, 1.16) P=0.862
At least 2000m	0.89 (0.75, 1.06) P=0.206	1.03 (0.78, 1.36) P=0.840	1.00 (0.88, 1.13) P=0.968	1.03 (0.90, 1.20) P=0.643
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace) (ref)_	1.00	1.00	1.00	1.00
Q2	1.00 (0.89, 1.13) P=0.965	0.94 (0.77, 1.15) P=0.528	1.03 (0.91, 1.16) P=0.675	0.97 (0.84, 1.11) P=0.630
Q3	0.92 (0.79, 1.08) P=0.310	1.08 (0.84, 1.39) P=0.556	1.00 (0.88, 1.14) P=0.987	1.02 (0.88, 1.17) P=0.805
Q4 (most greenspace)	0.76 (0.59, 0.97) P=0.031	1.17 (0.82, 1.67) P=0.388	1.04 (0.90, 1.21) P=0.593	1.02 (0.86, 1.19) P=0.850

4. Sensitivity analyses: Models additionally adjusted for baseline BMI, hypertension and medications for hypertension and high cholesterol

Supplementary Table 23. Modification of the association between built environment variables and hospital admissions due to CVD, by household income and area deprivation (adjusted for additional risk factors)

CVD-related admissions	*Annual household income*		*Area deprivation*	
	Less than £31,000	At least £31,000	More deprived	Less deprived
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.99 (0.93, 1.06) P=0.730	1.00 (ref)	0.91 (0.85, 0.99) P=0.028
One	1.00 (0.95, 1.08) P=0.793	0.94 (0.87, 1.01) P=0.100	0.93 (0.85, 1.03) P=0.191	0.91 (0.84, 0.99) P=0.036
2-3	1.04 (0.98, 1.11) P=0.140	0.99 (0.93, 1.07) P=0.998	0.99 (0.91, 1.09) P=0.912	0.96 (0.88, 1.04) P=0.305
4 or more	1.00 (0.97, 1.10) P=0.371	0.92 (0.86, 1.00) P=0.046	0.95 (0.87, 1.04) P=0.240	0.93 (0.85, 1.01) P=0.087
Stratum-specific HRs (4+ facilities vs 0)	1.02 (0.96, 1.10) P=0.479	0.94 (0.87, 1.02) P=0.140	0.93 (0.84, 1.03) P=0.149	1.02 (0.96, 1.09) P=0.531
Relative excess risk due to interaction (RERI)	-0.090 (-0.182, 0.002) P=0.055		0.066 (-0.030, 0.163) P=0.177	
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.91 (0.84, 0.99) P=0.031	1.00 (ref)	1.00 (0.92, 1.08) P=0.957
500-999m	0.97 (0.91, 1.03) P=0.343	0.91 (0.84, 0.98) P=0.009	1.01 (0.93, 1.09) P=0.849	0.95 (0.88, 1.02) P=0.133
1000-1999m	0.97 (0.91, 1.03) P=0.346	0.91 (0.84, 0.98) P=0.013	0.97 (0.89, 1.06) P=0.500	0.96 (0.89, 1.03) P=0.260
At least 2000m	0.93 (0.87, 1.00) P=0.050	0.92 (0.85, 1.00) P=0.049	1.02 (0.92, 1.13) P=0.657	0.93 (0.86, 1.00) P=0.064
Stratum-specific HRs (≥2000m vs <500m)	0.93 (0.86, 1.00) P=0.051	1.01 (0.92, 1.12) P=0.767	1.04 (0.93, 1.16) P=0.498	0.92 (0.86, 0.99) P=0.030
Relative excess risk due to interaction (RERI)	0.081 (-0.015, 0.177) P=0.098		-0.092 (-0.214, 0.030) P=0.141	
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.91 (0.84, 0.98) P=0.015	1.00 (ref)	0.99 (0.92, 1.07) P=0.879
Q2	0.99 (0.93, 1.05) P=0.640	0.93 (0.86, 1.00) P=0.058	1.01 (0.94, 1.09) P=0.778	0.96 (0.90, 1.02) P=0.197
Q3	0.99 (0.92, 1.06) P=0.714	0.94 (0.87, 1.01) P=0.089	1.03 (0.94, 1.13) P=0.489	0.96 (0.89, 1.02) P=0.171
Q4 (most greenspace)	0.97 (0.90, 1.05) P=0.515	0.95 (0.88, 1.04) P=0.266	0.97 (0.85, 1.10) P=0.611	0.97 (0.90, 1.04) P=0.401
Stratum-specific HRs (Q4 vs Q1)	0.99 (0.91, 1.09) P=0.883	1.02 (0.91, 1.13) P=0.774	0.96 (0.82, 1.12) P=0.596	0.97 (0.89, 1.05) P=0.409
Relative excess risk due to interaction (RERI)	0.071 (-0.023, 0.165) P=0.141		0.006 (-0.129, 0.141) P=0.934	

Supplementary Table 24. Modification of the association between neighbourhood environment variables and hospital admissions due to cancer, by household income and area deprivation (adjusted for additional risk factors)

Cancer-related admissions	*Annual household income*		*Area deprivation*	
	< £31,000	**At least £31,000**	**More deprived**	**Less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	0.96 (0.90, 1.02) P=0.224	1.00 (ref)	1.03 (0.95, 1.11) P=0.530
One	0.98 (0.92, 1.05) P=0.594	0.99 (0.92, 1.06) P=0.728	1.08 (0.97, 1.20) P=0.145	1.01 (0.93, 1.10) P=0.854
2-3	1.00 (0.94, 1.06) P=0.939	0.94 (0.88, 1.01) P=0.094	1.06 (0.96, 1.16) P=0.238	1.00 (0.92, 1.09) P=0.979
4 or more	0.99 (0.93, 1.05) P=0.695	0.90 (0.83, 0.97) P=0.004	1.01 (0.92, 1.11) P=0.775	0.98 (0.90, 1.07) P=0.722
Stratum-specific HRs (4+ facilities vs 0)	0.97 (0.91, 1.04) P=0.397	0.96 (0.88, 1.04) P=0.277	1.02 (0.93, 1.13) P=0.643	0.96 (0.90, 1.02) P=0.152
Relative excess risk due to interaction (RERI)	-0.051 (-0.139, 0.038) P=0.261		-0.056 (-0.165, 0.053) P=0.316	
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	0.89 (0.82, 0.97) P=0.007	1.00 (ref)	0.92 (0.85, 0.99) P=0.028
500-999m	0.93 (0.88, 1.00) P=0.035	0.93 (0.86, 1.00) P=0.046	0.90 (0.83, 0.97) P=0.006	0.94 (0.87, 1.01) P=0.076
1000-1999m	0.97 (0.91, 1.04) P=0.392	0.91 (0.84, 0.98) P=0.010	0.98 (0.90, 1.07) P=0.685	0.92 (0.85, 0.99) P=0.019
At least 2000m	0.94 (0.87, 1.01) P=0.076	0.91 (0.84, 0.98) P=0.019	0.92 (0.83, 1.03) P=0.137	0.91 (0.85, 0.99) P=0.020
Stratum-specific HRs (≥2000m vs <500m)	0.97 (0.90, 1.04) P=0.369	0.97 (0.88, 1.07) P=0.547	0.92 (0.82, 1.03) P=0.150	1.00 (0.93, 1.07) P=0.927
Relative excess risk due to interaction (RERI)	0.082 (-0.013, 0.176) P=0.090		0.073 (-0.039, 0.185) P=0.200	
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	0.86 (0.80, 0.93) P=0.000	1.00 (ref)	0.94 (0.87, 1.02) P=0.118
Q2	0.98 (0.92, 1.04) P=0.464	0.91 (0.84, 0.98) P=0.011	0.99 (0.92, 1.07) P=0.783	0.95 (0.89, 1.02) P=0.135
Q3	0.96 (0.89, 1.02) P=0.192	0.94 (0.87, 1.02) P=0.122	1.00 (0.91, 1.10) P=0.985	0.95 (0.89, 1.01) P=0.122
Q4 (most greenspace)	0.94 (0.87, 1.02) P=0.132	0.96 (0.88, 1.04) P=0.285	0.85 (0.75, 0.97) P=0.019	0.97 (0.90, 1.04) P=0.390
Stratum-specific HRs (Q4 vs Q1)	0.96 (0.88, 1.05) P=0.392	1.05 (0.95, 1.17) P=0.339	0.83 (0.71, 0.97) P=0.021	1.04 (0.95, 1.12) P=0.393
Relative excess risk due to interaction (RERI)	0.151 (0.062, 0.241) P=0.001		0.172 (0.045, 0.298) P=0.008	

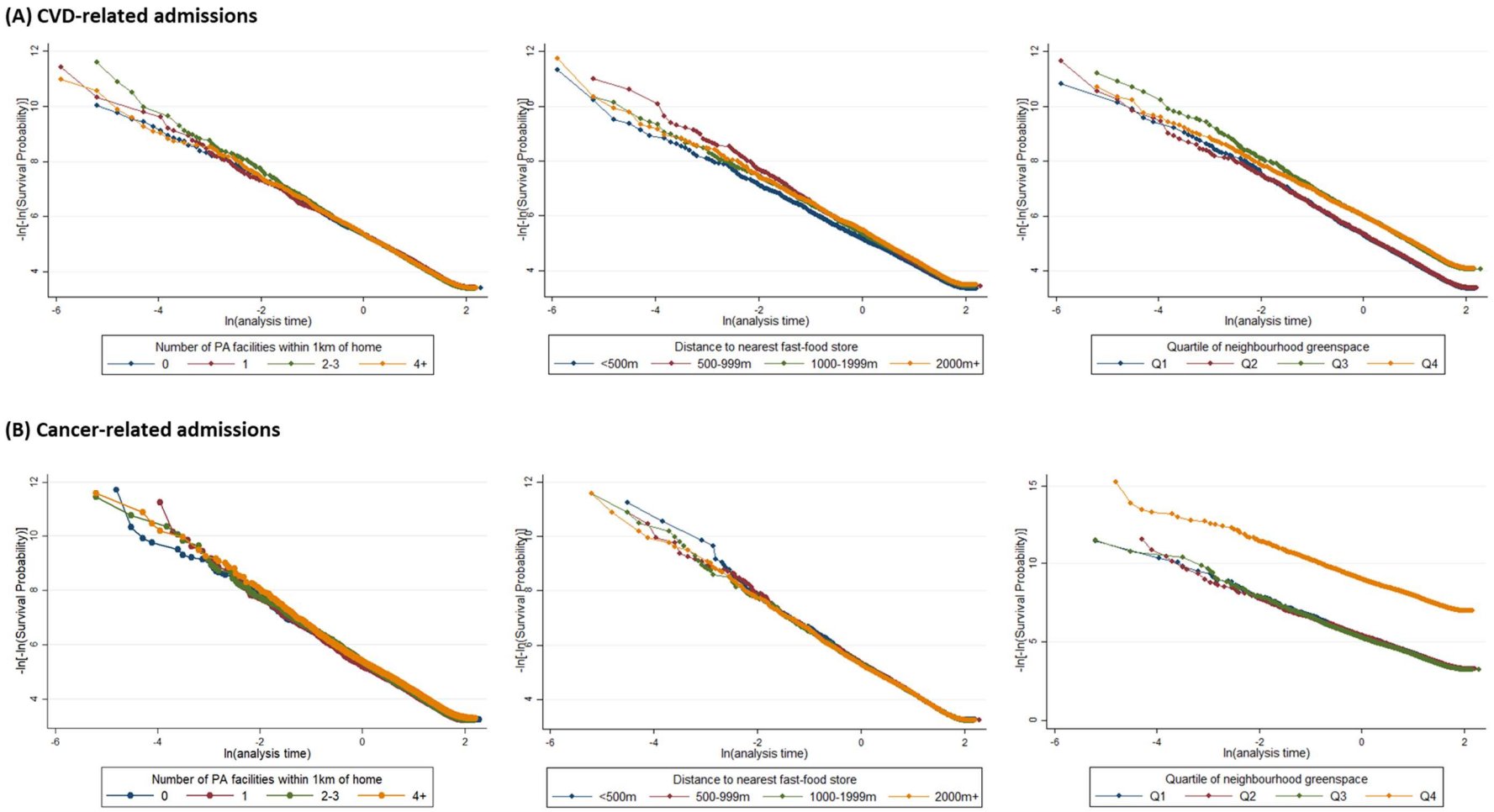
Supplementary Table 25. Association between neighbourhood characteristics and CVD-related hospital admissions, stratified by household income and area deprivation in combination (adjusted for additional risk factors)

CVD-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
One	0.97 (0.86, 1.09) P=0.629	0.82 (0.66, 1.00) P=0.050	1.03 (0.95, 1.11) P=0.527	0.97 (0.89, 1.05) P=0.468
2-3	1.01 (0.90, 1.13) P=0.880	0.91 (0.76, 1.09) P=0.293	1.07 (0.99, 1.15) P=0.109	1.02 (0.94, 1.11) P=0.665
4 or more	0.99 (0.88, 1.11) P=0.846	0.79 (0.65, 0.95) P=0.011	1.04 (0.96, 1.14) P=0.330	0.98 (0.89, 1.07) P=0.645
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
500-999m	1.02 (0.93, 1.12) P=0.614	1.00 (0.86, 1.16) P=0.991	0.93 (0.85, 1.01) P=0.101	0.98 (0.88, 1.09) P=0.732
1000-1999m	0.96 (0.87, 1.07) P=0.467	1.05 (0.88, 1.24) P=0.587	0.96 (0.88, 1.05) P=0.366	0.97 (0.87, 1.08) P=0.557
At least 2000m	1.00 (0.88, 1.14) P=0.998	1.16 (0.94, 1.44) P=0.171	0.89 (0.81, 0.98) P=0.018	0.98 (0.88, 1.10) P=0.767
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Q2	1.00 (0.91, 1.10) P=0.956	1.10 (0.95, 1.28) P=0.217	0.99 (0.90, 1.08) P=0.753	0.95 (0.85, 1.07) P=0.409
Q3	1.04 (0.93, 1.17) P=0.479	1.05 (0.86, 1.29) P=0.632	0.98 (0.88, 1.08) P=0.618	0.97 (0.87, 1.09) P=0.622
Q4 (most greenspace)	0.92 (0.77, 1.11) P=0.381	1.10 (0.83, 1.47) P=0.499	0.99 (0.89, 1.11) P=0.913	0.98 (0.87, 1.12) P=0.807

Supplementary Table 26. Association between neighbourhood characteristics and cancer-related hospital admissions, stratified by household income and area deprivation in combination (adjusted for additional risk factors)

Cancer-related admissions	*Combined household income and area deprivation*			
	Less than £31,000 & more deprived	**At least £31,000 & more deprived**	**Less than £31,000 & less deprived**	**At least £31,000 & less deprived**
Number of PA facilities	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
None	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
One	1.06 (0.94, 1.19) P=0.377	1.13 (0.92, 1.39) P=0.246	0.95 (0.88, 1.02) P=0.182	1.02 (0.94, 1.11) P=0.589
2-3	1.07 (0.96, 1.20) P=0.228	1.04 (0.85, 1.26) P=0.712	0.96 (0.88, 1.03) P=0.252	0.99 (0.91, 1.07) P=0.786
4 or more	1.05 (0.93, 1.18) P=0.403	0.98 (0.80, 1.19) P=0.804	0.94 (0.86, 1.02) P=0.156	0.97 (0.88, 1.06) P=0.449
Fast-food proximity	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Closer than 500m	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
500-999m	0.88 (0.80, 0.97) P=0.009	0.93 (0.80, 1.08) P=0.352	1.00 (0.91, 1.09) P=0.996	1.06 (0.96, 1.18) P=0.249
1000-1999m	0.96 (0.86, 1.06) P=0.418	1.01 (0.86, 1.20) P=0.880	1.03 (0.94, 1.12) P=0.558	0.98 (0.89, 1.09) P=0.742
At least 2000m	0.87 (0.76, 1.00) P=0.046	1.06 (0.85, 1.32) P=0.597	1.02 (0.93, 1.12) P=0.681	0.98 (0.88, 1.09) P=0.702
Greenspace	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*	*HR (95% CI)*
Q1 (least greenspace)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Q2	0.99 (0.90, 1.08) P=0.778	0.96 (0.82, 1.13) P=0.641	1.01 (0.92, 1.11) P=0.859	1.03 (0.92, 1.15) P=0.617
Q3	0.96 (0.85, 1.08) P=0.484	1.04 (0.85, 1.27) P=0.729	1.00 (0.91, 1.10) P=0.983	1.05 (0.94, 1.18) P=0.365
Q4 (most greenspace)	0.75 (0.62, 0.91) P=0.003	1.06 (0.79, 1.41) P=0.702	1.02 (0.91, 1.15) P=0.687	1.08 (0.95, 1.22) P=0.234

5. Examination of proportional hazards assumption



Supplementary Figure 15. Log-log plots (adjusted for all covariates) for graphical examination of proportional hazards assumption